# Relative Entropy, Distortion, the Bootstrap and Risk

by

R. Mark Reesor

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Statistics

Waterloo, Ontario, Canada, 2001

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

# Abstract

This thesis studies three related topics important to risk management, finance and insurance. The relative entropy bootstrap option pricing models are simulated pricing models where asset price movements are drawn from a set of historically observed movements. In particular, relative entropy is used to select the distribution that is closest to the empirical distribution and satisfies some prescribed moment conditions, such as the martingale constraint on the discounted asset price. A calibration algorithm for simulated pricing models is developed. The algorithm accommodates a set of options with different moneyness as well as across maturities. Finally, we establish the link between relative entropy and distortion. This connection gives a new perspective to the growing body of literature in which the Choquet integral is used to measure and price risks.

# Acknowledgements

It has been a pleasure to work under the supervision of Dr. Don McLeish and Dr. Adam Kolkiewicz. I wish to extend my sincere gratitude for their endless patience, encouragement and guidance during this endeavour. I also wish to thank my thesis committee: Dr. Hans Gerber, Dr. Phelim Boyle, Dr. Mary Hardy and Dr. Harry Panjer for their encouragement, helpful suggestions and time spent reviewing my work. I would like to recognise Dr. William Welch and Dr. Jock Mackay for their helpful advice and many illuminating discussions integral to my personal development and in the writing of this thesis.

My thanks also go out to the Natural Sciences and Engineering Research Council of Canada (NSERC), the Ontario Graduate Scholarship Program (OGS), Manulife Financial, Dr. Don McLeish, Dr. Adam Kolkiewicz, Dr. William Welch and Dr. Jerry Lawless for their generous financial support.

To my family and friends, thank you for your moral support and encouragement that was instrumental in completing this thesis and in helping me maintain a good quality of life during this process. I am forever indebted to my parents for their never-ending love, advice, and support, without which none of this would have been possible. Thank you.

Finally, to my wife and best friend Tonya, thank you. As with everything, I share this accomplishment with you and consider it as much yours as mine. You are the light of my life.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Overview

This thesis has evolved into a study of three related topics of importance in finance, risk management and insurance. The relative entropy bootstrap (REB) option pricing models are simulated pricing models in which asset price movements are drawn from a set of historically observed movements. While developing these simulated pricing models, the problem of calibration kept surfacing. This led to the creation of an algorithm to calibrate general simulated pricing models. Along the way, the similarity of minimum relative entropy distributions and distorted probability distributions became apparent. A formal link between distortion and minimum relative entropy as methods for re-weighting probability distributions has been established. This link provides new insight into the construction and testing of risk measures.

The rest of this section provides a brief summary of each of the three topics

while the remainder of this chapter gives motivation for developing the REB option pricing models. Some basic concepts in financial economics are given, along with a discussion of option pricing in incomplete markets. Section 1.4 introduces entropy methods, used extensively in Chapters 2 and 4, and gives a short summary of its use in financial applications. A discrete-time Markov market model is established in Section 1.5, providing the framework in which to cast the REB models of Chapter 2. This chapter concludes with a brief explanation of the bootstrap, which has different meaning in finance than it does in statistics. Chapter 2 contains the development and application of the REB pricing models. The calibration algorithm for simulated pricing models is discussed in Chapter 3. Following this, Chapter 4 establishes the connection between relative entropy and distortion. Application of this link to problems in finance, risk management and insurance are discussed.

### 1.1.1 The Relative Entropy Bootstrap

These are pricing models in which asset price movements are simulated from a set of historically observed movements. Relative entropy is used as a tool to adjust the empirical distribution of the observed price movements to allow risk-neutral valuation. These models generalise a nonparametric option pricing model proposed by Stutzer (1996). Desirable features of asset price movements such as fat tails and skewness are automatically incorporated into the model, provided that these features have been observed in the past. By simulating from the observed increments, the REB-I model provides a way to model asset price movements on the original scale and in such a manner as to preserve the discrete price jumps. These issues

are typically ignored by many pricing models.

We show that REB models are more consistent with the market prices of traded options than both the Black-Scholes and Stutzer models. Since price paths are generated during the simulation, REB models can value European securities with arbitrary payoff functions (e.g., barrier options). Furthermore, stochastic interest rate, dividend and volatility models can be accommodated, providing another attractive aspect of the REB models.

### 1.1.2 Calibration Algorithm for Simulated Pricing Models

Option pricing models typically have a parameter $\theta$ that governs the behaviour of the underlying stochastic variables. To produce reasonable prices, the model must be calibrated to the market by determining a value for $\theta$ that minimises, for example, the sum of squared pricing errors (model price minus market price) for a set of options. Models having a closed-form solution such as Black-Scholes are easily calibrated by taking derivatives and using standard gradient-based optimisation procedures. However, derivatives are unavailable for simulated models and, due to the simulation noise, model prices are not exactly known. These facts exclude many optimisation schemes.

Here, we approximate the pricing error locally with a polynomial function of $\theta$. The local polynomial is used to compute and optimise the sum of squared pricing errors. In order to fit the approximating polynomial, simulations must be performed at a number of different parameter settings. Clearly the computational expense increases with the number of settings. As such, the number of settings should

kept to a minimum while still providing enough information for a reasonable fit of the local polynomial. We use methods from experimental design to select these parameter settings. These methods are well-known in statistics and have the ability to maximise the information obtained in a fixed number of experimental runs. To the best of this author's knowledge, this is the first application of statistically designed experiments to the problem of calibrating simulated pricing models.

Furthermore, the algorithm can handle a set of options with different moneyness as well as across maturities. Examples using two well-known pricing models, along with the REB models developed here illustrate the success of the algorithm.

### 1.1.3 Relative Entropy, Distortion and Risk

Relative entropy and distortion are two ways of re-weighting probability distributions while maintaining the support set of the original distribution. Theorem 4.2 formally establishes the link between relative entropy and distortion. This link provides additional intuition behind the many premium principles and risk measures having a Choquet integral representation. Relative entropy provides an easy way of constructing new risk measures with a Choquet integral representation through the set of moment constraints. We provide conditions on the moment constraints, equivalently the distortion function, that ensure the constructed risk measure satisfies certain properties such as coherence. Furthermore, results are given that allow one to verify some properties of a given risk measure having a Choquet integral representation.

The connection between relative entropy and distortion gives a new perspec-

tive to the growing body of literature in which the Choquet integral is applied to problems in finance, risk management and insurance.

## 1.2 Motivation for the REB

Traditional approaches to option pricing are typically three-step processes. First, one assumes a particular stochastic process that drives the underlying asset price. This process specifies the dynamics of the asset under the objective or physical measure ($P$). Then, this process is transformed into a risk-neutral process, either using arguments about dynamic hedging, or by imposing assumptions about the market price of risk. The adjusted process governs asset prices under the risk-neutral measure ($Q$). The parameters of the process are estimated, giving a set of prices of options (Jackwerth (1999)). Intuitively, the process under the $P$ measure and the risk-neutral process should be similar, and historically observed prices should appear to have been generated by the posited process.

There have been many models proposed to describe the behaviour of assets in the $P$ measure. By far the most popular, partly because of its tractability, is a stochastic differential equation with geometric Brownian Motion. Here, the returns on a stock are assumed to be lognormally distributed. Adjustment of this process to the risk-neutral measure is accomplished through dynamic hedging requirements. The celebrated Black-Scholes model for European options is derived in this manner (Black and Scholes (1973)).

However, empirical evidence reveals some fairly serious problems with the log-normal assumption. One of the problems is that the distribution of observed returns

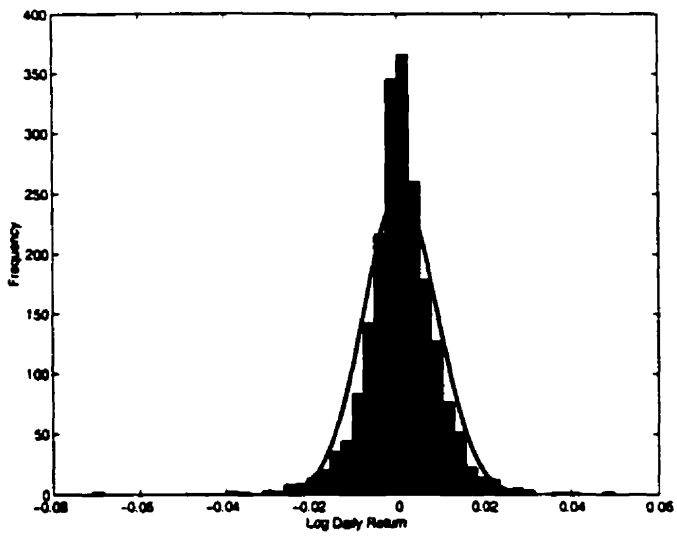


Figure 1.1: Histogram of daily logreturns on the S&P500 (March 15, 1991 to May 10, 1999), with fitted normal density.

typically has fatter tails than that assumed by the lognormal distribution. In other words, returns of large magnitude occur more often in the real world than can be explained by this type of model. Real returns have also been known to exhibit more skewness than is possible under this model. Examples of these problems are seen in Figure 1.1 which shows a histogram of the daily logreturns of the S&P500 index from March 15, 1991 to May 10, 1999. A normal density fit to this data is superimposed on the histogram, clearly showing that the returns are not lognormally distributed. Similar evidence that the returns are not lognormal is seen in Figure 1.2, which is a histogram of the daily logreturns on Royal Bank of Canada shares.

A multitude of stochastic volatility models, geometric Brownian motion subordinated to a random clock and other processes such as jump diffusions and gamma

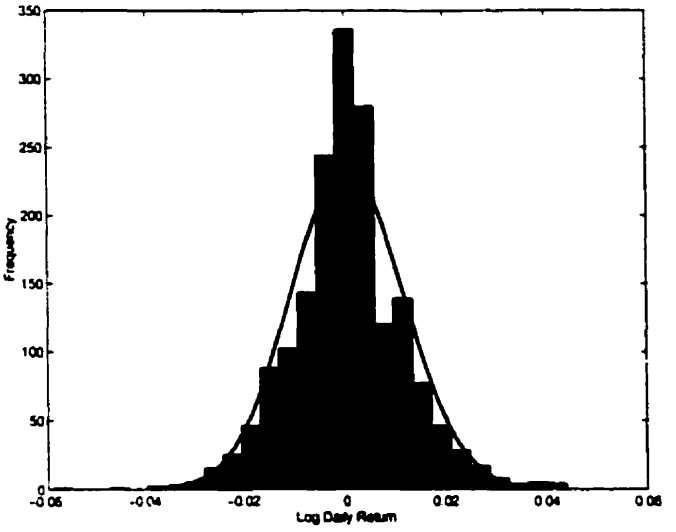


Figure 1.2: Histogram of daily logreturns on the Royal Bank shares (Nov. 11, 1990 to Dec. 31, 1996), with fitted normal density.

processes are some approaches that have been used to address these deficiencies with varying degrees of success. Merton (1976), Cox and Ross (1976), Jarrow and Rudd (1982), Engle (1982), Rubinstein (1985), Bollerslev (1986), Wiggins (1987), Hull and White (1987), Stein and Stein (1991), Heston (1993), Heston (1993), Amin and Ng (1993), Duan (1995), Redekop (1995), Duan (1997), Bakshi, Cao and Chen (1997), McLeish (1998), Hardy (1999) and Ritchken and Trevor (1999) are but some of the references in the literature to the specification, implementation and evaluation of these alternative models. Most of the above models impose some sort of continuous, parametric assumptions on the underlying stock price process.

Another problem with the lognormal approach is the assumption that asset returns are continuous. This problem, and its relevance to option pricing, was noted in Merton (1976). Since returns are computed from the asset price, the

assumption that returns are continuous is a consequence of the assumption that the asset price is continuous. In fact, this is not a valid assumption. Some exchanges quote stock prices to the nearest $\frac{1}{16}^{\text{th}}$ or $\frac{1}{32}^{\text{nd}}$ of a dollar while others quote prices to the nearest cent. Nevertheless, prices are contained on a discrete lattice of points which fail to produce a continuum of possible returns. Numerical difficulties are sometimes experienced when trying to estimate continuous models from discrete data (Hasbrouck (1999)).

Furthermore, several studies have shown evidence of *price clustering*, the tendency for prices to fall more frequently on certain values than on others, enhancing the effect of the discrete nature of prices (see Chapter 3 of Campbell et al (1997)). Reasons for these values being favoured over others will not be investigated here. Price discreteness and tick size are important in the setting of the bid/ask spread, as discussed by Hasbrouck (1999).

As an illustration of price discreteness and clustering, consider the daily closing value, $S_t$, of the Royal Bank of Canada shares from November 11, 1990 to December 31, 1996. There are over 1500 observed daily increments, $\Delta S_t = S_t - S_{t-1}$, yet only 76 *unique values* are present. Figure 1.3 is a histogram showing the frequency of these unique increments, with each value as its own bin (each bar is centred over a unique increment value). Notice the large spikes that correspond to observed daily changes in the stock price of $0, \pm\frac{1}{8}^{\text{th}}, \pm\frac{2}{8}^{\text{th}}, \pm\frac{3}{8}^{\text{th}}$ and $\pm\frac{4}{8}^{\text{th}}$ of a dollar.

Stock prices and options written on stocks are not the only financial instruments in which price discreteness is an issue. In fact, the prices of all instruments are inherently discrete. Interest rates are quoted to the nearest basis point (1 basis

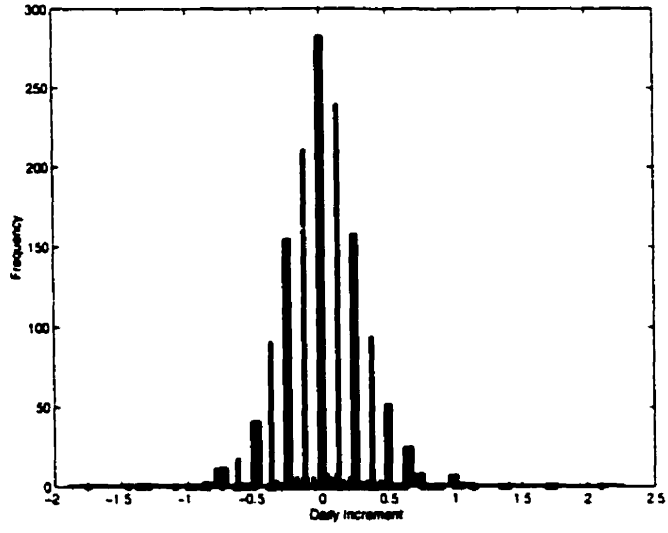


Figure 1.3: Histogram of daily increments of Royal Bank Shares (Nov. 11, 1990 to Dec. 31, 1996) with each unique increment as the centre of its own bin.

point=0.01%) and the Cad/US exchange rate to the nearest hundredth of a cent. Commodity futures are sometimes quoted to the nearest $\frac{1}{4}$ cent per unit (along with the number of units in the contract size) or to the nearest hundredth of a cent per unit. For example, on the New York Mercantile Exchange (NYM) each Heating Oil No. 2 futures contract is for 42,000 gallons and the price is quoted to the nearest ten thousandth of a dollar per gallon (Hull (1997)). The price of this contract must then be a multiple of \$4.20.

The problem of price discreteness is not unique to the lognormal model, but to any continuous-state model of asset price returns (or asset prices). Essentially all diffusion models and all other models assuming a continuous-state stochastic process for the asset returns (prices) such as the gamma model for stock option pricing (Gulko (1998)) and the shifted gamma and shifted inverse gaussian models

(Gerber and Shiu (1994)) have this shortcoming. However, it is unclear if price discreteness has a major impact on the pricing and hedging of derivatives and this issue may be more of theoretical interest than practical importance.

The option pricing models proposed in Chapter 2 do not make any parametric distributional assumptions about the stock price dynamics. By using the empirical distribution of observed price movements, these models capture the observed skewness and kurtosis. The REB-I model specifically addresses the issue of price discreteness by simulating from the observed daily increments. On the other hand, REB-L1 and REB-L2 ignore the issue of price discreteness as they simulate movements from the observed logreturns. In addition, they have computational advantages over REB-I and are more attractive from the perspective of stationarity.

## 1.3 Basic Principles of Finance

In this section we review some well-known definitions, concepts and results of modern financial theory. Further results and (formal) definitions can be found in Duffie (1996), Panjer et al (1998) and Björk (1998). We begin with the definition of arbitrage, and the usual assumption regarding arbitrage and the market in a discrete-time setting.

**Definition 1.1 (Arbitrage).** An *arbitrage* is an opportunity to make riskless profits in the market.

**Assumption 1.1 (No Arbitrage Opportunities).** The market does not admit arbitrage opportunities, or if it does, the opportunities disappear immediately.

**Definition 1.2 (Self-financing Trading Strategy).** A *portfolio* or *trading strategy*, $\theta_t$, represents the portfolio held after trading at time $t = 0, 1, \ldots, T$. A trading strategy is *self-financing* if there are no exogenous infusions or withdrawals of money.

**Definition 1.3 (Reachable Contingent Claim).** A *contingent claim* with maturity date $T$ is random variable whose value is known at time $T$. A contingent claim is *reachable* if it can be reproduced by some self-financing trading strategy. This strategy is called the *replicating* or *hedging* portfolio.

**Definition 1.4 (Complete Market).** A market model is *complete* if all contingent claims are reachable, otherwise it is *incomplete*.

### Complete Markets

In a discrete multiperiod market, suppose there is a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_t\}$ defined on $(\Omega, \mathcal{F}, P)$. Also, suppose there is an asset (possibly a vector) with price process $\{S_t\}$ and a risk-free bond with price process $\{B_t\}$ such that $S_t$ and $B_t$ are $\mathcal{F}_t$-measurable for all $t$. Furthermore, trading takes place only at the discrete times $\{0, 1, 2, \ldots, T\}$ and there are no transaction costs. We have the following well-known result.

**Theorem 1.1 (EMM).** *If the market is complete and Assumption 1.1 holds, then there exists a unique measure $Q$ such that for all $t < T$*

$$E^Q\left[\frac{S_T}{B_T}\middle|\mathcal{F}_t\right] = \frac{S_t}{B_t}. \tag{1.1}$$

*This unique measure $Q$ is called the* Equivalent Martingale Measure, EMM, *as $Q$ is equivalent to the measure $P$ and the normalised price process $\{\frac{S_t}{B_t}\}$ is a $Q$ martingale.*

The measure $Q$ is also known as the *risk-neutral measure.*

### Incomplete Markets

Suppose we have the same setup as above. Then, the following result for incomplete markets is analogous to Theorem 1.1 for complete markets, with the notable absence of uniqueness of the measure $Q$.

**Theorem 1.2 (EMM).** *If the market is incomplete and Assumption 1.1 holds, then there exists a measure $Q$ such that for all $t < T$*

$$E^Q\left[\frac{S_T}{B_T}\middle|\mathcal{F}_t\right] = \frac{S_t}{B_t}, \tag{1.2}$$

*where $Q$ is equivalent to $P$.*

With appropriate changes to the definitions, there are corresponding results to Theorems 1.1 and 1.2 in continous-time market models.

## 1.3.1 Specifying a Risk-neutral Measure

In complete markets, the risk-neutral measure is uniquely specified. However, in incomplete markets, Theorem 1.2 asserts only the existence of an equivalent martingale measure, $Q$. There may be many measures equivalent to $P$ that satisfy Equation 1.2. Typically the market may impose more constraints than Equation

1.2 on the measure, effectively shrinking the set of martingale measures, $\mathcal{M}$, from which to choose an appropriate pricing measure. These additional constraints may be the market prices of options on the underlying asset, the prices of futures contracts on the underlying asset, or a combination of both.

These market constraints can be written as moment conditions on a martingale measure. Consider the inverse problem of determining the distribution, $Q$, of a random variable, $X$, satisfying

$$\int_{\Omega} G_i(X)\mathrm{d}Q = E^Q[G_i(X)] = C_i, \qquad \text{for} \quad 1 \leq i \leq N, \tag{1.3}$$

where $G_1, \ldots, G_N$ are given functions and $C_1, \ldots, C_N$ are given numbers. Here, $X$ is a state variable describing the economy and $C_i$ is the current market price of a traded security with future cash flow $G_i(X)$, for $1 \leq i \leq N$ (Avellaneda (1998)). The $C_i$'s could also be the calibrating values of certain parameters governing the evolution of $X$. Clearly the constraint in Equation 1.2 can be rewritten in the form of Equation 1.3. The constraints are used to *calibrate* the model to the market (i.e., the pricing measure should be able to reproduce the market prices of the traded instruments).

The problem of pricing assets in incomplete markets can be broken down into two components:

1. What constraints should be imposed on the set of equivalent martingale measures?

2. Given the constraints, how should one select a particular measure to price

the asset?

A paper by Chernov and Ghysels (1998) provides information about the first component. Many methods have been proposed to address the second question.

One approach is to specify risk preferences through a utility function and then select the measure that maximises the expected utility of terminal wealth (Karatzas et al (1991), Rogers (1994), and Karatzas and Shreve (1999)). Another tactic is to select a measure that minimises some risk function (Föllmer and Sondermann (1986), Föllmer and Schweizer (1991), Föllmer and Leukert (1998), Colwell and Elliot (1993), and Carr et al (2000)). One can also select an EMM by choosing the one that is closest to the physical measure $P$ while satisfying the required moment constraints. Minimum relative entropy is a natural tool to perform this type of measure selection (Frittelli (2000)). Related to minimum relative entropy are conditional Esscher transforms which are discussed by Bühlmann et al (1996). Of course there are connections among all of these methods as noted by Bühlmann et al (1996), Samperi (1998), Karatzas and Shreve (1999), Frittelli (2000), and Rouge and El Karoui (2000), for example.

The pricing models developed here use minimum relative entropy to select a risk-neutral measure. Also, the link between relative entropy and distortion in Chapter 4 reveals that distortion methods can also be used to select a pricing measure. Using its relationship to relative entropy, it may be interesting to investigate how distortion relates to other methods of measure selection.

### 1.3.2 The Inverse Problem

Reverse financial engineering is another approach of determining a risk-neutral measure by solving an inverse problem. Specifically, given a set of market prices of options, a natural question to pose is if these option prices contain any information about the distribution or stochastic process driving the price of the underlying asset. Obviously, only those distributions/processes that are compatible with the observed option prices are plausible.

Breeden and Litzenberger (1978) were the first to show that market prices of options contain information about the conditional (risk-neutral) distribution of the underlying asset. In particular, for European options expiring at time $T$, they show that

$$q_T(K) = e^{rT}\frac{\partial^2 C}{\partial K^2}, \qquad K > 0, \tag{1.4}$$

where $q_T$ is the conditional risk-neutral density, $C(K,T)$ is the market price of the option with strike $K$, and $e^{rT}$ is a interest factor. Thus, knowledge of the option prices over all possible strikes is enough to completely determine the pricing measure (in complete markets). This work was extended by Dupire (1994) and Derman and Kani (1994) who showed that the risk-neutral diffusion process can be inferred from a complete set of option prices (all strikes and all maturities). Bick (1982) extends the Breeden and Litzenberger result to the case where the underlying stock returns are discontinuous.

Although theoretically appealing, these results depend on a continuum of option prices and complete markets. Of course, no real market is complete, nor does it

have a continuum of option prices. Recent papers by Derman and Kani (1998), Ledoit and Santa-Clara (1998), and Britten-Jones and Neuberger (2000) extend the above results to include stochastic volatility, although their methods still require a complete set of options.

There have been many recent papers proposing solutions to the inverse problem. Jackwerth (1999) gives a survey on these various methods and classifies them into two main categories—parametric and nonparametric. Parametric methods are divided into three groups. The first group is expansion methods which start with a basic distribution and add correction terms to make it more flexible. An example of this is a recent article by Rubinstein (1998) who starts with a binomial model and adds correction terms using Edgeworth expansions.

The second group consists of parametric distributions that are more flexible than the lognormal distribution. The gamma distribution, different Burr distributions and the Johnson family of distributions have all been used (Jackwerth (1999)). The third parametric group involves mixture models in which a flexible distribution is created from mixing standard distributions. Melick and Thomas (1997) use a mixture of three lognormal distributions to imply the risk-neutral distribution of crude oil futures from American option prices.

Nonparametric methods are also split up into three categories. Kernel estimators, such as that used by Aït-Sahalia and Lo (1998) make up the first category. Entropy methods make up the second group. These will be discussed more fully in the next section.

The other group is called curve-fitting by Jackwerth. Generally, these meth-

ods try to fit implied volatilities or the risk-neutral probability density as well as possible with some flexible function. Methods for fitting the volatility smile can be found in Jackwerth (1999). The fitted risk-neutral density is typically the one that minimises the distance to a prior distribution, subject to pricing the options correctly. Rubinstein (1994) and Jackwerth and Rubinstein (1996) are two papers advocating this approach. In the latter, the authors show that the implied risk-neutral distribution is robust to the choice of distance measure.

## 1.4 Entropy Methods

### 1.4.1 Maximum Entropy

Entropy is a measure of the flatness and smoothness of a distribution. It can be thought of as a measure of uncertainty of a random variable $X$ having distribution $P$. The higher the entropy, the more uncertainty there is about the random variable and its distribution is more spread out. Given a finite number of states, entropy is maximised when $P$ is a discrete uniform distribution (maximum uncertainty) and it is minimised when one of the states occurs with probability one (minimum uncertainty).

**Definition 1.5 (Entropy).** The *entropy* of a discrete probability measure $P$ having support set $\mathcal{X}$ is defined as

$$H(P) = -\sum_{k \in \mathcal{X}} p_k \ln p_k, \tag{1.5}$$

where $p_k = P(\text{state } k)$ for each $k \in \mathcal{X}$. The entropy of a discrete random variable $X$, $H(X)$, is defined using the distribution induced by $X$.

*Remark 1.1.* It is sometimes useful to think of entropy as an expected value by rewriting Equation 1.5 as

$$E^P\left[\ln \frac{1}{Y}\right] = -\sum_{k \in \mathcal{X}} p_k \ln p_k = H(P), \tag{1.6}$$

where $Y$ is a random variable defined by $P(Y = p_k) = p_k$ for all $k \in \mathcal{X}$.

*Remark 1.2.* There is a corresponding definition of entropy when $P$ is a continuous probability measure. Specifically, if $g$ is the density function of a random variable $X$ with distribution $P$, then the *differential entropy* is

$$H(P) = -\int_{\Omega} g(x) \ln g(x) dx, \tag{1.7}$$

where $\Omega$ is the support of $X$ (e. g., the real line) (Cover and Thomas (1991), Kapur and Kesavan (1992)).

Entropy maximisation is a commonly used approach to solve under-specified inverse problems not just in finance, but in many other areas of science. Suppose that we have a discrete state space, $\Omega$, and some moment conditions like Equation 1.3. The state space and the moment constraints together form the information set, $\mathcal{I}$. This could be an under-specified problem, as there may be many distributions defined on $\Omega$ that satisfy the constraints. Entropy maximisation selects a unique distribution consistent with $\mathcal{I}$ in the most unbiased way. That is, of all the distributions consistent with $\mathcal{I}$, the distribution that maximises entropy is the one that

is least committal with respect to unknown or missing information. In this sense there is no reason to prefer any other distribution over the maximum entropy distribution (MED) (Samperi (1998), Cover and Thomas (1991), Gulko (1998), Buchen and Kelly (1996)). Essentially, this is the maximum entropy principle as stated in Kapur and Kesavan (1992).

The MED, $P^*$, solves a convex optimisation problem. Specifically, given a discrete state space with $m$ possible states and a random variable $X$ which is in one of the $m$ possible states, $P^*$ solves

$$\begin{aligned} \max_P H(P) &= \max_P \left\{ -\sum_{i=1}^{m} p_i \ln p_i \right\} \qquad (1.8) \\ \text{subject to} \quad \sum_{i=1}^{m} p_i &= 1, \\ E^P[G_j(X)] &= C_j, \qquad \text{for} \quad 1 \le j \le N, \\ p_i &> 0 \text{ for all } i, \end{aligned}$$

where $G_1, \ldots, G_N$ and $C_1, \ldots, C_N$ are defined as in Equation 1.3. Note that the first constraint combined with the positivity constraints force $P^*$ to be a distribution. This optimisation problem is easily solved by Lagrangian methods.

**Theorem 1.3.** *If there is at least one measure satisfying the constraints, then the unique solution $P^*$ to the constrained maximum entropy optimisation problem is*

$$p_i^* = \exp\left( \sum_{j=1}^{N} \lambda_j G_j(x_i) - \psi(\lambda_1, \ldots, \lambda_N) \right) \qquad \text{for } i = 1, \ldots, m, \qquad (1.9)$$

*where*

$$\psi(\lambda_1, \ldots, \lambda_N) = \ln \left\{ \sum_{i=1}^{m} \exp \left( \sum_{j=1}^{N} \lambda_j G_j(x_i) \right) \right\} \tag{1.10}$$

*and $\psi$ satisfies*

$$\frac{\partial \psi}{\partial \lambda_j} = C_j \qquad for\ j = 1, \ldots, N. \tag{1.11}$$

*(Cover and Thomas (1991), Kapur and Kesavan (1992)).*

A result analogous to Theorem 1.3 also holds in the case where $P$ is a continuous distribution (Gulko (1998), Cover and Thomas (1991), Kapur and Kesavan (1992)).

### 1.4.2 Minimum Relative Entropy

While entropy measures the amount of uncertainty about a probability distribution, relative entropy provides a measure of distance from one probability distribution to another. Relative entropy is also known as a measure of directed divergence from one distribution to another. There are many different versions of relative entropy. The Kullback-Leibler measure is by far the most popular and, unless otherwise specified will be taken as the measure of relative entropy used throughout this thesis. The following subsection gives a family of generalised measures of directed divergence that is used in Chapter 4.

**Definition 1.6 (Relative Entropy).** Let $P$ and $Q$ be two discrete probability

measures. The *relative entropy* or *cross-entropy* of $Q$ with respect to $P$ is

$$\begin{aligned} H(Q|P) &= \sum q_k \ln \frac{q_k}{p_k}, \qquad \text{if} \qquad Q \ll P, \\ &= +\infty, \qquad\qquad\qquad \text{otherwise,} \end{aligned} \tag{1.12}$$

where $Q \ll P$ means that $Q$ is absolutely continuous with respect to $P$. The relative entropy of a random variable $Y$ with respect to the random variable $X$, $H(Y|X)$, is defined using the distributions induced by $Y$ and $X$. Note that $H(Q|P) < \infty$ implies $Q \ll P$.

*Remark 1.3.* There is a corresponding definition of relative entropy when $P$ and $Q$ are continuous probability measures. Specifically,

$$\begin{aligned} H(Q|P) &= \int_\Omega \mathrm{d}Q \ln \left\{ \frac{\mathrm{d}Q}{\mathrm{d}P} \right\}, \qquad \text{if} \qquad Q \ll P, \\ &= +\infty, \qquad\qquad\qquad \text{otherwise,} \end{aligned} \tag{1.13}$$

where $\frac{\mathrm{d}Q}{\mathrm{d}P}$ is the Radon-Nikodym derivative of $P$ with respect to $Q$.

*Remark 1.4.* In the special case where the prior distribution $P$ is uniform (discrete or continuous), then minimising the relative entropy of $Q$ with respect to $P$ is equivalent to maximising the entropy of $Q$.

*Remark 1.5.* As in Remark 1.1, it is may be useful to think of relative entropy as an expected value by rewriting it as

$$E^Q \left[ \ln \left\{ \frac{\mathrm{d}Q}{\mathrm{d}P} \right\} \right] = H(Q|P), \tag{1.14}$$

when $Q \ll P$ and with the obvious interpretation when $Q$ and $P$ are discrete measures.

*Remark 1.6.* Although $H(Q|P) \geq 0$, with equality if and only if $Q = P$, relative entropy is not a metric since it is asymmetric in its arguments. At times, it may be convenient to work with the symmetric version

$$H(Q, P) = H(Q|P) + H(P|Q). \tag{1.15}$$

Note that $H(Q, P) < \infty$ implies $Q \sim P$ ($Q$ and $P$ have the same null sets) (for further information, see Cover and Thomas (1991), Samperi (1998), or Kapur and Kesavan (1992)).

In Section 1.4.1 we found the MED, $P^*$, consistent with some information set $\mathcal{I}$. This set contained only the state space, $\Omega$, and the $N$ moment constraints. It did not include any other information about the distribution, $P$, of the random variable $X$. Typically, we may have some information or may want to make assumptions about this distribution. Then, intuitively we would like to find the distribution, $Q$, satisfying the $N$ moment constraints that is as *close* to $P$ as possible. This is the principle of minimum relative entropy given in Kapur and Kesavan (1992).

We measure closeness with relative entropy, $H(Q|P)$, and it can be thought of as the information lost by using $Q$ in place of $P$. We would like to minimise this information loss by finding the distribution, $Q^*$, that minimises $H(Q|P)$.

The minimum relative entropy distribution (MRED), $Q^*$, solves a convex optimisation problem. Specifically, given a discrete state space with $m$ possible states and a random variable $X$ which is in one of the $m$ possible states with *prior* dis-

tribution, $P$, $Q^*$ solves

$$\begin{aligned} \min_Q H(Q|P) &= \min_Q \left\{ \sum_{i=1}^{m} q_i \ln \frac{q_i}{p_i} \right\} \qquad (1.16) \\ \text{subject to} \quad \sum_{i=1}^{m} q_i &= 1, \\ E^Q[G_j(X)] &= C_j, \qquad \text{for} \quad 1 \le j \le N, \\ q_i &\ge 0 \text{ for all } i, \end{aligned}$$

where $G_1, \ldots, G_N$ and $C_1, \ldots, C_N$ are defined as in Equation 1.3. Note that the first constraint combined with the nonnegativity constraints force $Q^*$ to be a distribution. This optimisation problem is easily solved by Lagrangian methods.

**Theorem 1.4.** *If there is a measure satisfying the constraints, then the unique solution, $Q^*$, to the constrained minimum relative entropy optimisation problem is*

$$q_i^* = p_i \exp\left( \sum_{j=1}^{N} \lambda_j G_j(x_i) - \psi(\lambda_1, \ldots, \lambda_N) \right) \qquad \text{for } i = 1, \ldots, m, \qquad (1.17)$$

*where*

$$\psi(\lambda_1, \ldots, \lambda_N) = \ln \left\{ \sum_{i=1}^{m} p_i \exp\left( \sum_{j=1}^{N} \lambda_j G_j(x_i) \right) \right\} \qquad (1.18)$$

*and $\psi$ satisfies*

$$\frac{\partial \psi}{\partial \lambda_j} = C_j \qquad \text{for } j = 1, \ldots, N. \qquad (1.19)$$

*(Cover and Thomas (1991), Kapur and Kesavan (1992)).*

A result analogous to Theorem 1.4 is also available in the continuous case (Samperi (1998), Cover and Thomas (1991), Buchen and Kelly (1996), Kapur and Kesavan (1992)).

In the field of information theory, the distribution $Q^*$ is also known as the *minimum cross entropy distribution* (Buchen and Kelly (1996), Cover and Thomas (1991), Jackwerth and Rubinstein (1996)). While in statistics, $Q^*$ is the *exponential tilt of* $P$ (Efron and Tibshirani (1993)) and in actuarial science, it is the *Esscher transform of* $P$ (Gerber and Shiu (1994), Panjer et al (1998)).

### 1.4.3 Generalised Entropy

As mentioned above, there are many measures of directed divergence. Kapur and Kesavan (1992) suggest six properties that a measure of directed divergence may possibly be required to satisfy. These are:

1. $\tilde{H}(Q|P) \geq 0$,
2. $\tilde{H}(Q|P) = 0$ if and only if $Q = P$,
3. $\tilde{H}(Q|P) = H(P|Q)$,
4. $\tilde{H}(Q|P) + \tilde{H}(P|R) \geq \tilde{H}(Q|R)$,
5. $\tilde{H}(Q|P)$ is a convex function of $Q$, and
6. when $\tilde{H}(Q|P)$ is minimised subject to some given linear constraints using Lagrange's method, the minimising probabilities should be $\geq 0$.

In order to provide a measure of distance it is essential that $\tilde{H}$ satisfies properties 1 and 2. Properties 3 and 4 are desirable as these are characteristics of a metric, while properties 5 and 6 are wanted for mathematical convenience. The Kullback-Leibler measure satisfies properties 1,2,5 and 6.

**Definition 1.7 (Generalised Measure of Relative Entropy).** Measures satisfying properties 1,2, and 5 are called *generalised measures of relative entropy* or *directed divergence.*

**Definition 1.8 (Csiszer's Directed Divergence).** For continuous distributions, Csiszer's family of measures of directed divergence is given by

$$H^C(Q|P) = \int_\Omega \phi\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right)\mathrm{d}P. \tag{1.20}$$

where $\phi$ is a twice-differentiable convex function for which $\phi(1) = 0$.

Csiszer's measure satisfies properties 1,2 and 5 for all appropriate $\phi$, and sometimes satisfies property 6 (Kapur and Kesavan (1992) Section 7.2). The Kullback-Leibler measure is a special case of this family with $\phi(x) = x\ln x$. In Section 4.5.2 we use the generalised entropy optimisation principle stated below to define a generalised measure of directed divergence given an a priori distribution, an a posteriori distribution and a set of moment constraints. We restrict our search of a generalised measure to Csiszer's family of directed divergences and determine an appropriate function $\phi$.

*Remark 1.7 (The Generalised Entropy Optimisation Principle).* Given any three of the following four probabilistic entities:

1. a priori probability distribution $P$,

2. a posteriori probability distribution $Q$,

3. the set of moment constraints $C$, and

4. the measure of entropy or relative entropy $M$,

the fourth should be chosen such that either the entropy of $Q$ is maximum or the relative entropy of $Q$ relative to $P$ is minimum (Kapur and Kesavan (1992) p.298).

### 1.4.4 Entropy Applications in Finance

There are many recent examples in the financial literature making use of the concept of entropy. In his PhD thesis, Gulko develops the Entropy Pricing Theory (EPT), which gives an economic justification for using MED's to price options. In his framework, he shows that the EPT includes the usual Arbitrage Pricing Theory (APT) and also derives entropic versions of the Black-Scholes and capital asset pricing models (Gulko (1998)). Minimum relative entropy also has an economic justification, as shown by Samperi (1998), who develops an equivalence relation between utility maximisation and minimising the relative entropy to a prior distribution. Gerber and Shiu (1994) derive alternative European option pricing models using various prior distributions and the Esscher transform.

The use of entropy as a calibration or inverse problem solving tool is the area in which most of the applications have appeared. Buchen and Kelly (1996) show the MED accurately estimates a known distribution given simulated option prices

at different strikes. In the same article, they also show that the MRED accurately estimates a known distribution given simulated option prices at different strikes.

Given a set of market option prices over a range of strikes and expiration dates, Samperi (1998), Avellaneda et al (1997), Avellaneda (1998), and Laurent and Leisen (1999), use minimum relative entropy to infer the risk-neutral dynamics of the underlying asset consistent with these market prices. None of the above applications make use of historical information of the underlying asset in defining the prior distribution (diffusions and trinomial tree models are used). These methods rely on a lot of current market information and use entropy to fill in the gaps between the observed prices.

Recently, Avellaneda et al (2001) use minimum relative entropy to calibrate Monte Carlo asset pricing models. They mention using historical information on the underlying asset, such as rates of return and volatility, as well as current market information of liquid instruments. However, they use Gaussian random shocks in their simulation and they make no mention of the issue of price discreteness. As with the above methods, this method is designed to use a lot of current market information.

Stutzer (1996) uses a nonparametric model for stock returns calibrated using minimum relative entropy to price European stock options. In particular, he uses the historical distribution of asset price returns to construct a prior distribution. Options are then priced using the MRED with the only constraint that the discounted underlying asset is a martingale. He advocates this approach over other methods as it does not impose a parametric model on the underlying asset and it

does not require market prices covering a wide range of strikes. This is a predictive model for pricing options, rather than interpolating the price of a new option between strikes and across expiration dates. An approach similar to this is developed for pricing bond futures options (Stutzer and Chowdhury (1999)). Drawbacks to Stutzer's approach are that the prior distribution of returns is rather restrictive and no attempt is made to address the issue of price discreteness.

Other examples of entropy methods in finance include an alternative to generalised method of moments (GMM) estimation proposed by Kitamura and Stutzer (1997) and a method of testing conditional distributions in dynamic models proposed by Duan and So (1999). Entropy methods have also been used for some time in insurance applications (Berliner and Lev (1980), Maeder (1982) and Sundt (1982)).

## 1.5 Markov Market Model

The setup used here is an extended version of that used by Laurent and Leisen (1999). Here, we use a generic state variable to describe the evolution of the market, while they used only the discounted asset price (they noted the extension to a more general state variable in their article).

Suppose there is a finite set of states, $\mathcal{S} = \{s_1, \ldots, s_m\}$, a discrete set of dates, $\mathcal{T} = \{0, \ldots, T\}$, a state variable, $\{X_t\}_{t \in \mathcal{T}}$, and a sequence $\{\Sigma_t\}_{t \in \mathcal{T}}$ where $\Sigma_t \subset \mathcal{S}$ denotes the set of possible states of $X_t$ and $N_t$ is the cardinality of $\Sigma_t$ for each $t \in \mathcal{T}$. Furthermore, suppose there is a sequence of matrices $\{\Pi_t\}_{t \in \mathcal{T} \setminus T}$ such that for each $t$, $\Pi_t$ is a stochastic matrix on $\Sigma_t \times \Sigma_{t+1}$ of dimension $N_t \times N_{t+1}$ (i.e., the

elements of $\Pi_t$ are nonnegative and the rows sum to one). The notation $\mathcal{T} \setminus T$ denotes the set of times $\{0, 1, \ldots, T-1\}$ excluding time $T$.

Define the set of possible paths as $\Omega = \otimes_{t \in \mathcal{T}} \Sigma_t$, let $\mathcal{F}$ be the set of all subsets of $\Omega$, and define a filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ on $\Omega$ as $\mathcal{F}_t = \sigma\{X_0, \ldots, X_t\}$ for each $t \in \mathcal{T}$. Suppose $\{X_t\}_{t \in \mathcal{T}}$ is a Markov Process. Then, the evolution of $\{X_t\}_{t \in \mathcal{T}}$ on $\Omega$ is described by the probability measure $P_k$ defined by

$$P_k(X_{t+1} = \bar{s}_{t+1} | X_t = \bar{s}_t) = \Pi_t(\bar{s}_t, \bar{s}_{t+1}), \tag{1.21}$$

for all states $\bar{s}_t \in \Sigma_t$, $\bar{s}_{t+1} \in \Sigma_{t+1}$, for each $t \in \mathcal{T} \setminus T$, and where $k$ corresponds to some initial state in $\mathcal{S}$. By definition, $\Sigma_0 = k$ and $\Pi_0$ has only one row.

A simple result of the Markovian assumption, is that the probability of any sample path $X = (X_0, X_1, \ldots, X_T) \in \Omega$ can be computed as

$$P_k(X = x) = P_k(X_0 = \bar{s}_0, X_1 = \bar{s}_1, \ldots, X_T = \bar{s}_T) = \prod_{t=0}^{T-1} \Pi_t(\bar{s}_t, \bar{s}_{t+1}). \tag{1.22}$$

Note that a condition for $X$ to be in $\Omega$ is for $X_0 = k$. Also, for any times $\tau_1, \tau_2 \in \mathcal{T}$ with $\tau_1 < \tau_2$, we can compute a matrix of transition probabilities between dates $\tau_1$ and $\tau_2$ by

$$\Pi_{\tau_1, \tau_2} = \prod_{u=\tau_1}^{\tau_2 - 1} \Pi_u. \tag{1.23}$$

We now have a well-defined probability space, $(\Omega, \mathcal{F}, P_k)$, a filtration, $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ defined on $\Omega$, and sequences $\{\Sigma_t\}_{t \in \mathcal{T}}$ and $\{\Pi_t\}_{t \in \mathcal{T} \setminus T}$ describing the sets of possible

states and the 1-step transition probabilities at each time. Note that by the construction of $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$, $X_t$ is $\mathcal{F}_t$-measurable for all $t \in \mathcal{T}$. The spirit of the Markov assumption is made apparent in the Lemma below (similar to Lemma 3C in Duffie (1996)).

**Lemma 1.5.** *Assume that $\{X_t\}_{t\in\mathcal{T}}$ is a Markov process. For any $t \in \mathcal{T}$ define an arbitrary function $f : \otimes_{u=t}^{T}\Sigma_u \mapsto \mathcal{R}$ where $\mathcal{R}$ is the range of $f$. Then there exists a fixed function $g_k : \Sigma_t \mapsto \mathcal{R}$ such that for any initial state $k \in \mathcal{S}$,*

$$E^{P_k}[f(X_t,\ldots,X_T)|\mathcal{F}_t] = E^{P_k}[f(X_t,\ldots,X_T)|X_t] = g_k(X_t) \qquad (1.24)$$

*where $E^{P_k}$ denotes expectation under $P_k$.*

To show how Lemma 1.5 applies to modeling of asset prices, consider the situation in which the state variable $X_t$ is the price of a risky asset at time $t$. On $\otimes_{u=t}^{T}\Sigma_u$ define the function $f(X_t,\ldots,X_T) = X_T$, the price of the asset at time $T$. With this definition, the range of $f$ is a subset of $\Sigma_T$. Now given the initial price of the asset $X_0 = k$, we have

$$E^{P_k}[f(X_t,\ldots,X_T)|\mathcal{F}_t] = E^{P_k}[X_T|X_t] = g_k(X_t), \qquad (1.25)$$

for some function $g_k$. Under Assumption 1.1, Equation 1.25 can be converted to a more useful form by using a risk-free bond and an equivalent martingale measure.

### 1.5.1 Risk-neutral Measure for a Markovian Price Process

Suppose that we have the same setup as above and that the state variable $X_t = S_t$, the asset price at time $t$. We are assuming the asset price process $\{S_t\}_{t\in\mathcal{T}}$ is a discrete-time Markov chain with initial value $S_0 = k$ whose evolution on $\Omega$ is described by the sequences $\{\Sigma_t\}_{t\in\mathcal{T}}$ and $\{\Pi_t\}_{t\in\mathcal{T}\setminus T}$. Furthermore, assume there is a risk-free bond worth $B_t$ at time $t$. Also, we make the usual assumptions of a frictionless market where trading occurs only at times $t = 0, 1, 2 \ldots, T$.

Under the assumption of no arbitrage opportunities, Theorem 1.2 asserts the existence of a risk-neutral measure, $Q_k$, such that

$$E^{Q_k}\left[\frac{S_{t+1}}{B_{t+1}}\middle|\mathcal{F}_t\right] = \frac{S_t}{B_t} \quad \forall\, t \in \mathcal{T}\setminus T.$$

As in Section 1.3.1, there may be additional (linearly independent) moment constraints imposed on $Q_k$ at each $t \in \mathcal{T}\setminus T$ namely,

$$E^{Q_k}[G_{m,t}(S_{t+1})|\mathcal{F}_t] = C_{m,t}, \qquad \text{for} \quad 2 \le m \le M_t, \tag{1.26}$$

where $G_{2,t}, \ldots, G_{M_t,t}$ are given $\mathcal{F}_{t+1}$-measurable functions and $C_{2,t}, \ldots, C_{M_t,t}$ are given $\mathcal{F}_t$-measurable functions.

Since $\{S_t\}_{t\in\mathcal{T}}$ is Markovian, Lemma 1.5 asserts that for each $t \in \mathcal{T}\setminus T$ the moment constraints can be expressed as

$$E^{Q_k}[G_{m,t}(S_{t+1})|S_t] = C_{m,t} \qquad \text{for} \quad 1 \le m \le M_t, \tag{1.27}$$

with the martingale constraint corresponding to $m = 1$.

We select our risk-neutral measure, $Q_k^*$, by taking the MRED that satisfies the required constraints and is absolutely continuous with respect to $P_k$. Namely, $Q_k^*$ solves

$$\begin{aligned} \min_{Q_k} H(Q_k|P_k) &= \min_{Q_k} \left\{ \sum_{s\in\Omega} Q_k(S=s) \ln \frac{Q_k(S=s)}{P_k(S=s)} \right\} \\ \text{subject to} \quad \sum_{s\in\Omega} Q_k(S=s) &= 1, \\ E^{Q_k}[G_{m,t}(S_{t+1})|S_t] &= C_{m,t}, \qquad \text{for} \quad 1 \le m \le M_t, \ \forall\, t, \\ Q_k &\ll P_k, \end{aligned} \tag{1.28}$$

where $G_{1,t}, \ldots, G_{M_t,t}$ and $C_{1,t}, \ldots, C_{M_t,t}$ are defined as in Equation 1.27.

If there is at least one measure satisfying the constraints in 1.28, then by Theorem 1.4 the unique solution, $Q_k^*$, to the above optimisation problem has transition matrices with elements

$$\Pi_t^*(s_t, s_{t+1}) = \Pi_t(s_t, s_{t+1}) \exp\left( \sum_{m=1}^{M_t} \lambda_{m,t} G_{m,t}(s_t) - \psi_t(\lambda_{1,t}, \ldots, \lambda_{M_t,t}) \right), \tag{1.29}$$

where

$$\psi_t(\lambda_{1,t}, \ldots, \lambda_{M_t,t}) = \ln \left\{ \sum_{s_{t+1}\in\Sigma_{t+1}} \Pi_t(s_t, s_{t+1}) \exp\left( \sum_{m=1}^{M_t} \lambda_{m,t} G_{m,t}(s_t) \right) \right\} \tag{1.30}$$

for all $s_t \in \Sigma_t$, $s_{t+1} \in \Sigma_{t+1}$ and $t \in \mathcal{T} \setminus T$. The function $\psi_t$ satisfies

$$\frac{\partial \psi_t}{\partial \lambda_{m,t}} = C_{m,t} \quad \text{for } m = 1, \ldots, M_t, \ \forall\, t. \tag{1.31}$$

Thus, the risk-neutral probability of any sample path $S = (S_0, S_1, \ldots, S_T) \in \Omega$ can be computed as

$$Q_k^*(S = s) = Q_k^*(S_0 = \bar{s}_0, S_1 = \bar{s}_1, \ldots, S_T = \bar{s}_T) = \prod_{t=0}^{T-1} \Pi_t^*(\bar{s}_t, \bar{s}_{t+1}). \tag{1.32}$$

Note that since $S \in \Omega$, this implies that $\bar{s}_0 = k$.

If there is no measure that satisfies the constraints, then one of the assumptions of the market model has been violated. There may be an arbitrage opportunity, or one or more inappropriate moment constraints were imposed on $\{S_t\}_{t \in \mathcal{T}}$. Assuming a frictionless market, or ignorance of the bid/ask spread of the asset price are two other possible explanations for the non-existence of $Q_k^*$. Also, the market may not be Markovian, or if it is Markovian, the correct state variable may not have been used (i.e., perhaps an extended stated variable of the asset price paired with the current volatility is needed). Theoretically, the state variable can be extended until the market is Markovian, leaving one (or more) of the other possible assumption violations as the reason for the non-existence of $Q_k^*$ (Campbell et al (1997)).

## 1.6 The Bootstrap

The nonparametric bootstrap is a resampling technique that allows researchers to analyse problems without making any distributional assumptions about the data. There is also a parametric bootstrap which assumes some distributional form. The bootstrap is computationally intensive but, as modern computers evolve, its use is becoming increasingly popular. Here, we will primarily be concerned with the nonparametric bootstrap.

Often, it is of interest to estimate characteristics (e.g., mean, variance, quantiles) of some unknown distribution $F$. These quantities can be thought of as functionals of the distribution function, $\theta(F)$. Consider independent and identically distributed (i.i.d.) observations $x_1, \ldots, x_n$ from an unknown distribution $F$. Then, an estimate of $\theta(F)$ is given by $\theta(\hat{F})$ where $\hat{F}$ is the empirical distribution function of $F$.

Now, consider drawing a simple random sample with replacement $(X_1^*, \ldots, X_{m_n}^*)$ from $\hat{F}$. This is called a *bootstrap sample.* The functional of interest can then be estimated using the bootstrap sample, denoted $\hat{\theta}^*(\hat{F})$. If we draw $B$ independent bootstrap samples and average the estimate from each sample, we obtain the bootstrap estimate of the functional,

$$\hat{\theta}^*(\hat{F}) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^*(\hat{F}) , \tag{1.33}$$

where $\hat{\theta}_b^*(\hat{F})$ is the estimate of $\theta(\hat{F})$ for the $bth$ bootstrap sample. The standard error of the bootstrap estimate is easily obtained by calculating the standard deviation of the $\hat{\theta}_b^*(\hat{F})'s$ and dividing by $\sqrt{B}$. Pointwise confidence intervals can be

constructed in the obvious manner (Efron (1983)). Asymptotic properties of bootstrap estimators, such as consistency, have been extensively explored and are fairly well-known (see Efron and Tibshirani (1993) for more information).

# Chapter 2

# The Relative Entropy Bootstrap (REB)

## 2.1 Introduction

In most of this chapter, we use the value of a stock (or index), $S_t$, as the state variable. We ignore the fact that the assumed discrete-time market model is violated by trading between the time points. Related to this point is the *nonsynchronous trading effect* which occurs when time series are taken to be recorded at time intervals of one length when, in fact, they are recorded at time intervals of other, possibly irregular lengths (see Chapter 3.1 of Campbell et al (1997)). For example, the daily closing value is the value at which the last reported trade occurred on that business day, and the actual times between successive closing values may be different than 24 hours. This effect is ignored here.

We use the discrete-time Markovian market model of Section 1.5 to develop

relative entropy bootstrap (REB) option pricing models based on the observed history of the underlying asset. The REB-I pricing model introduced in Section 2.2.1 attempts to preserve the observed price discreteness/clustering by simulating risk-neutral price movements from the observed increments. An alternative to this approach is to simulate risk-neutral price movements from the observed logreturns. Two bootstrap models, REB-L1 and REB-L2, use this approach and are developed in Section 2.2.2. The differences between REB-L1 and REB-L2 are due to moment constraints imposed directly on the stock price versus moment constraints placed on the logreturns. Section 2.3 discusses the differences between these approaches, in the context of the Black-Scholes (lognormal) model and the REB models. The REB pricing models are generalisations of a nonparametric pricing model proposed by Stutzer (1996) which is briefly described in Section 2.4. This is followed by a discussion of the calibration of pricing models to the market prices of options. Section 2.6 provides a short description of pricing options and estimating hedging parameters via simulation.

As an illustration, these models along with the Black-Scholes model are used to price S&P500 index call options and are compared with the observed market prices. Furthermore, hedging parameters for the REB models are estimated and compared to the corresponding Black-Scholes hedging parameters. This section ends with a discussion of the benefits and drawbacks of the REB models.

The remainder of the chapter investigates two possible reasons for the pricing error pattern observed in the S&P500 example. Section 2.8 gives the results of a simulation study, showing that the pricing error pattern observed for the S&P500

example can be obtained when the underlying data generating process is the Hull-White stochastic volatility model (Hull and White (1987)). The chapter ends with a section investigating the assumption of independence of the logreturns in the risk-neutral world. Using the observed S&P500 call option prices, implied risk-neutral distributions are computed at a number of different maturities. These implied distributions clearly show that any model for the risk-neutral evolution of the underlying asset must have some sort of dependency or a jump component of sufficient persistence that it overrides the central limit theorem (CLT) in the medium term. If it does not, then the pricing model will not be consistent with the full set of option prices.

## 2.2 REB

In this section, three models are presented that we call relative entropy bootstrap models. These models are dynamic versions of the model proposed by Stutzer (1996) that is described in Section 2.4. The first model simulates future stock price movements by randomly drawing increments from the set of observed increments. This model will be referred to as REB-I. The other two models simulate future stock price moves by randomly drawing logreturns from the set of observed logreturns. In one model, moment constraints are placed on the stock price, while in the other model, moment constraints are prescribed on the logreturns. These models shall be referred to as REB-L1 and REB-L2 respectively.

In Section 2.3 we explore the differences between REB-L1 and REB-L2, showing that, for the maturities considered, there is essentially no difference between them.

We also compare these two models with REB-I and notice some difference in the cdfs particularly at the shortest maturity studied.

### 2.2.1 REB-I

Suppose we have a fairly long history of closing prices, $s_0, \ldots, s_N$, with many "ups" and "downs" at many different levels of the price (i.e., a dataset where it is possible to get estimates of the $P$ transition probabilities). From this history, we compute the observed increments, $y_n = s_n - s_{n-1}$ for $n = 1, \ldots, N$, and then construct the set of unique observed increments $\mathcal{X}$. Corresponding to each $x \in \mathcal{X}$ is the observed relative frequency, $w_x$, of that increment. That is,

$$w_x = \frac{\sum_{n=1}^{N} I(y_n = x)}{N} \quad \text{for all} \quad x \in \mathcal{X}, \tag{2.1}$$

where $I(A) = 1$ if $A$ is true and is zero otherwise.

Let the initial stock price be $k$ and consider the somewhat naive approach of using a random walk on the observed increments to model the dynamics of the stock price under the measure $P_k$ (see Equation 2.3). In other words, given $S_t$,

$$S_{t+1} = S_t + X_{t+1}, \tag{2.2}$$

where $X_{t+1}$ is an increment drawn from $\mathcal{X}$ according to $P_k$ ($X_{t+1}$ is not to be confused with the state variable).

Using the notation of Section 1.5 and $\Sigma_0 = k$, we construct the set of possible prices at $t = 1$ by $\Sigma_1 = \Sigma_0 \oplus \mathcal{X}$, where $\oplus$ denotes adding each element of $\Sigma_0$ with

each element of $\mathcal{X}$. Similarly, $\Sigma_2 = \Sigma_1 \oplus \mathcal{X}$ and we continue in this fashion until $\Sigma_T = \Sigma_{T-1} \oplus \mathcal{X}$. Then, the set of all possible paths is defined as $\Omega = \otimes_{t \in \mathcal{T}} \Sigma_t$, the $\sigma$-algebra $\mathcal{F}$ is the set of all possible subsets of $\Omega$ and the filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ is defined by $\mathcal{F}_t = \sigma(S_0, \ldots, S_t)$ for all $t \in \mathcal{T}$.

Assuming the stock price follows a random walk on the observed increments, the evolution of $\{S_t\}_{t \in \mathcal{T}}$ on $\Omega$ is described by the probability measure $P_k$ defined by

$$\begin{aligned} P_k(S_{t+1} = \bar{s}_{t+1} | S_t = \bar{s}_t) &= \Pi_t(\bar{s}_t, \bar{s}_{t+1}) \\ &= P_k(X_{t+1} = \bar{s}_{t+1} - \bar{s}_t | S_t = \bar{s}_t) \\ &= P_k(X_{t+1} = x | S_t = \bar{s}_t) \end{aligned} \tag{2.3}$$

for all states $\bar{s}_t \in \Sigma_t, \bar{s}_{t+1} \in \Sigma_{t+1}$, for each $t \in \mathcal{T} \setminus T$ and for some increment $x \in \mathcal{X}$. From the Markov property of the random walk, the probability of any path $S = (S_0, S_1, \ldots, S_T) \in \Omega$ can be computed as

$$P_k(S = s) = P_k(S_0 = \bar{s}_0, S_1 = \bar{s}_1, \ldots, S_T = \bar{s}_T) = \prod_{t=0}^{T-1} \Pi_t(\bar{s}_t, \bar{s}_{t+1}). \tag{2.4}$$

Given $S_0 = k$, suppose we want to simulate a stock price path to time $T$. At each $t \in \mathcal{T} \setminus T$, we need only a draw random increment, $X_{t+1}$, from $\mathcal{X}$ according to $\Pi_t$ and use Equation 2.2 to update the price of the stock. In order to carry out this simulation, we must specify the elements of $\Pi_t$. As we are assuming a random walk, it is natural to estimate these probabilities with the empirical distribution of

the increments, namely

$$\widehat{\Pi}_t(\bar{s}_t, \bar{s}_{t+1}) = \begin{cases} w_x & \text{if } \bar{s}_{t+1} = \bar{s}_t + x \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

for each $\bar{s}_t \in \Sigma_t, \bar{s}_{t+1} \in \Sigma_{t+1}$, for some $x \in \mathcal{X}$ and all $t \in \mathcal{T} \setminus T$. Let $\widehat{P}_k$ denote the estimated measure.

Under this specification, the simulated stock price paths are equivalently generated using a nonparametric bootstrap of the observed increments. However, these bootstrap paths are not generated under a risk-neutral measure. The results of Section 1.5.1 give a method to generate risk-neutral bootstrap price paths as described below.

### A Risk-neutral Bootstrap

Typical pricing models constrain the mean (drift) and the variance (volatility) of the stock price dynamics. The required mean is dictated by the no arbitrage assumption, imposing the martingale restriction on the discounted stock price, while a (stochastic) volatility model determines the required variance of the increments.

To generate a risk-neutral bootstrap, the idea is to reweight the sampling probabilities used to draw the random increment at each time point. These new weights should impose the required mean and variance on the random increment and be as *close* as possible to the observed relative frequencies. As mentioned in Chapter 1 relative entropy is used as the measure of closeness between two distributions.

Given $S_0 = k$, consider simulating a stock price path $S \in \Omega$ to time $T$ from the

observed increments. At each $t$, suppose the annualised risk-free interest rate from $t$ to $t+1$ (of length $\Delta t$) is $r_t$ and $\sigma_t$ is the required volatility of the increments. Then, we want the measure $Q_k^*$ that solves

$$\begin{aligned}
\min_{Q_k} H(Q_k|\widehat{P}_k) &= \min_{Q_k}\left\{\sum_{s\in\Omega} Q_k(S=s)\ln\frac{Q_k(S=s)}{\widehat{P}_k(S=s)}\right\} \\
\text{subject to} \quad \sum_{s\in\Omega} Q_k(S=s) &= 1, \\
E^{Q_k}[S_{t+1}|S_t] &= S_t e^{r_t\Delta t}, \quad \forall\, t\in\mathcal{T}\setminus T, \\
E^{Q_k}[S_{t+1}^2|S_t] &= S_t^2\sigma_t^2\Delta t + S_t^2 e^{2r_t\Delta t}, \quad \forall\, t\in\mathcal{T}\setminus T, \\
Q_k &\ll \widehat{P}_k. \qquad (2.6)
\end{aligned}$$

If there is at least one measure satisfying the constraints in 2.6, then by Theorem 1.4 the unique solution, $Q_k^*$, to the above optimisation problem has transition matrices with elements

$$\Pi_t^*(s_t, s_{t+1}) = \widehat{\Pi}_t(s_t, s_{t+1})\exp\left\{\lambda_{1,t}(s_{t+1}-s_t) + \lambda_{2,t}(s_{t+1}-s_t)^2 - \psi(\lambda_{1,t},\lambda_{2,t})\right\},$$

where

$$\psi(\lambda_{1,t},\lambda_{2,t}) = \ln\left\{\sum_{s_{t+1}\in\Sigma_{t+1}} \widehat{\Pi}_t(s_t, s_{t+1})\exp\left(\lambda_{1,t}(s_{t+1}-s_t) + \lambda_{2,t}(s_{t+1}-s_t)^2\right)\right\},$$

for all $s_t\in\Sigma_t$, $s_{t+1}\in\Sigma_{t+1}$ and $t\in\mathcal{T}\setminus T$. The Lagrange multipliers $\lambda_{1,t}$ and $\lambda_{2,t}$ are uniquely determined to satisfy the constraints for each $t\in\mathcal{T}\setminus T$. Therefore,

the risk-neutral probability of any sample path $S = (S_0, S_1, \ldots, S_T) \in \Omega$ is

$$Q_k^*(S = s) = Q_k^*(S_0 = \bar{s}_0, S_1 = \bar{s}_1, \ldots, S_T = \bar{s}_T) = \prod_{t=0}^{T-1} \Pi_t^*(\bar{s}_t, \bar{s}_{t+1}). \qquad (2.7)$$

The REB-I model has the advantage of modelling stock price movements in the original scale and preserves the observed price discreteness. However, the $P$ measure transition matrices ($\Pi_t$ or the estimates $\widehat{\Pi}_t$) do not reflect relative changes in the stock price (a possible stationarity issue). Relative price changes are reflected in the risk-neutral transition matrices, but these are computed to be close to their $P$ measure counterparts. Another drawback of this approach is a positive probability that the asset price will become negative, although we have not encountered this problem to date.

### 2.2.2 REB-Logreturns

Although REB-I preserves the observed price discreteness, an obvious criticism of this approach is that the empirical distribution of the increments does not reflect relative price changes. An alternative is to use a bootstrap on the observed log-returns. This approach helps to address the stationarity issue as the $P$ measure now accounts for relative price changes, but ignores the observed price discreteness. Another appealing feature of this approach over REB-I is that the possibility of negative asset prices is eliminated.

From the observed price history of daily closing prices, we compute the observed daily logreturns, $\tilde{r}_n = \ln\left(\frac{s_n}{s_{n-1}}\right)$, for $n = 1, \ldots, N$, and then construct the set of unique observed logreturns $\mathcal{R}$. Corresponding to each $r \in \mathcal{R}$ is the observed relative

frequency, $w_r$, of that logreturn. That is,

$$w_r = \frac{\sum_{n=1}^{N} I(\tilde{r}_n = r)}{N}, \quad \text{for all} \quad r \in \mathcal{R}. \tag{2.8}$$

Suppose that the initial stock price is $k$ and consider modelling the dynamics of the stock price under the measure $P_k$ (see Equation 2.10) as a random walk on the observed logreturns. In other words, given $S_t$,

$$S_{t+1} = S_t e^{R_{t+1}}, \tag{2.9}$$

where $R_{t+1} \in \mathcal{R}$ is a logreturn drawn from $\mathcal{R}$ according to $P_k$.

Using the notation of Section 1.5 and $\Sigma_0 = k$, the set of possible prices at $t = 1$ is constructed by $\Sigma_1 = \Sigma_0 \odot e(\mathcal{R})$, where $e(\mathcal{R})$ denotes element-wise exponentiation of the set $\mathcal{R}$ and $\odot$ denotes mulitplying each element of $\Sigma_0$ with each element of $e(\mathcal{R})$. Similarly, $\Sigma_2 = \Sigma_1 \odot e(\mathcal{R})$ and we continue in this fashion until $\Sigma_T = \Sigma_{T-1} \odot e(\mathcal{R})$. Then, the set of all possible paths is defined as $\Omega = \otimes_{t \in \mathcal{T}} \Sigma_t$, the $\sigma$-algebra $\mathcal{F}$ is the set of all possible subsets of $\Omega$ and the filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ is defined by $\mathcal{F}_t = \sigma(S_0, \ldots, S_t)$ for all $t \in \mathcal{T}$.

Assuming the stock price follows a random walk on the observed logreturns, the

evolution of $\{S_t\}_{t\in\mathcal{T}}$ on $\Omega$ is described by the probability measure $P_k$ defined by

$$\begin{aligned} P_k(S_{t+1} = \bar{s}_{t+1}|S_t = \bar{s}_t) &= \Pi_t(\bar{s}_t, \bar{s}_{t+1}) \\ &= P_k\left(\bar{s}_t e^{R_{t+1}} = \bar{s}_t e^r | S_t = \bar{s}_t\right) \\ &= P_k\left(R_{t+1} = r | S_t = \bar{s}_t\right) \\ &= P_k\left(R_{t+1} = r\right) \end{aligned} \tag{2.10}$$

for all states $\bar{s}_t \in \Sigma_t, \bar{s}_{t+1} \in \Sigma_{t+1}$, for each $t \in \mathcal{T}$ and for some logreturn $r \in \mathcal{R}$. From the Markov property of the random walk on the observed logreturns, the probability of any path $S = (S_0, S_1, \ldots, S_T) \in \Omega$ can be computed as

$$P_k(S = s) = P_k(S_0 = \bar{s}_0, S_1 = \bar{s}_1, \ldots, S_T = \bar{s}_T) = \prod_{t=0}^{T-1} \Pi_t(\bar{s}_t, \bar{s}_{t+1}). \tag{2.11}$$

Given $S_0 = k$, suppose we want to simulate a stock price path to time $T$. At each $t \in \mathcal{T} \setminus T$, we need only a draw random logreturn, $R_{t+1}$, from $\mathcal{R}$ according to $\Pi_t$ and use Equation 2.9 to update the price of the stock. In order to carry out this simulation, we must specify the elements of $\Pi_t$. Since we are assuming a random walk on the observed logreturns, it is natural to estimate these probabilities with the empirical distribution, namely

$$\widehat{\Pi}_t(\bar{s}_t, \bar{s}_{t+1}) = \begin{cases} w_r & \text{if } \bar{s}_{t+1} = \bar{s}_t e^r \\ 0 & \text{otherwise,} \end{cases} \tag{2.12}$$

for each $\bar{s}_t \in \Sigma_t, \bar{s}_{t+1} \in \Sigma_{t+1}$, for some $r \in \mathcal{R}$ and all $t \in \mathcal{T} \setminus T$.

Under this specification, the simulated stock price paths are equivalently generated using a nonparametric bootstrap of the observed logreturns. However, these bootstrap paths are not generated under a risk-neutral measure. As with REB-I, relative entropy minimisation is used to adjust the selection probabilities in order to simulate price paths under an appropriate pricing measure.

### REB-L1

For this model, moment constraints are imposed on the stock price $S_t$ and the risk-neutral transition probabilities are computed analagously to REB-I, with the obvious substitution of increments for logreturns.

Given $S_0 = k$, consider simulating a stock price path $S \in \Omega$ to time $T$ from the observed logreturns. At each $t$, suppose the annualised risk-free interest rate from $t$ to $t+1$ (of length $\Delta t$) is $r_t$ and $\sigma_t$ is the required volatility of $S_t$. Then, we want the measure $Q_k^*$ that solves

$$\begin{aligned} \min_{Q_k} H(Q_k|\widehat{P}_k) &= \min_{Q_k}\left\{\sum_{s\in\Omega} Q_k(S=s)\ln\frac{Q_k(S=s)}{\widehat{P}_k(S=s)}\right\} \\ \text{subject to} \quad \sum_{s\in\Omega} Q_k(S=s) &= 1, \\ E^{Q_k}[S_{t+1}|S_t] &= S_t e^{r_t\Delta t}, \quad \forall\, t \in \mathcal{T}\setminus T, \\ E^{Q_k}[S_{t+1}^2|S_t] &= S_t^2\sigma_t^2\Delta t + S_t^2 e^{2r_t\Delta t}, \quad \forall\, t \in \mathcal{T}\setminus T, \\ Q_k &\ll \widehat{P}_k. \end{aligned} \tag{2.13}$$

Note that these are the same moment constraints imposed in Equation 2.6. The difference between these two optimisation problems is due to the different prior

distributions $\widehat{P}_k$.

If there is at least one measure satisfying the constraints in 2.13, then by Theorem 1.4 the unique solution, $Q_k^*$, to the above optimisation problem has transition matrices with elements

$$\Pi_t^*(s_t, s_{t+1}) = \widehat{\Pi}_t(s_t, s_{t+1}) \exp\left\{\lambda_{1,t} s_{t+1} + \lambda_{2,t} s_{t+1}^2 - \psi(\lambda_{1,t}, \lambda_{2,t})\right\}, \tag{2.14}$$

where

$$\psi(\lambda_{1,t}, \lambda_{2,t}) = \ln\left\{\sum_{s_{t+1} \in \Sigma_{t+1}} \widehat{\Pi}_t(s_t, s_{t+1}) \exp\left(\lambda_{1,t} s_{t+1} + \lambda_{2,t} s_{t+1}^2\right)\right\}, \tag{2.15}$$

for all $s_t \in \Sigma_t$, $s_{t+1} \in \Sigma_{t+1}$ and $t \in \mathcal{T} \setminus T$. The Lagrange multipliers $\lambda_{1,t}$ and $\lambda_{2,t}$ are uniquely determined to satisfy the constraints at each $t \in \mathcal{T} \setminus T$. Therefore the risk-neutral probability of any sample path $S = (S_0, S_1, \ldots, S_T) \in \Omega$ is

$$Q_k^*(S = s) = Q_k^*(S_0 = \bar{s}_0, S_1 = \bar{s}_1, \ldots, S_T = \bar{s}_T) = \prod_{t=0}^{T-1} \Pi_t^*(\bar{s}_t, \bar{s}_{t+1}). \tag{2.16}$$

Although considerable effort has been spent casting REB-L1 in the Markov market model terminology, it can be described in a much simpler fashion. Equation 2.10 shows that the $P$ transition probabilities are independent of the current state. These transition probabilities are also independent of the initial state $S_0 = k$. Therefore, the moment constraints in 2.13 can be written unconditionally as con-

straints on the observed logreturns, namely,

$$\begin{aligned} E^Q[e^{R_{t+1}}] &= e^{r_t \Delta t}, \quad \forall\, t \in \mathcal{T} \setminus T, \\ E^Q[e^{2R_{t+1}}] &= \sigma_t^2 \Delta t + e^{2r_t \Delta t}, \quad \forall\, t \in \mathcal{T} \setminus T. \end{aligned} \tag{2.17}$$

This explicitly shows that REB-L1 is a nonparametric bootstrap of the observed logreturns, with the selection probabilities suitably re-weighted via minimum relative entropy.

### REB-L2

Here, moment constraints are placed on the logreturn, $R_t$, rather than the stock price as in REB-L1. We shall dispense with the notation of the Markov market model and note that the distribution of logreturns is independent of the stock price. The moment constraints we propose to use are consistent with the moments of the logreturns in the Black-Scholes model. Specifically, we replace the moment constraints in Equation 2.17 with

$$\begin{aligned} E^Q[R_{t+1}] &= \left(r_t - \frac{\sigma_t^2}{2}\right) \Delta t, \\ E^Q[R_{t+1}^2] &= \sigma_t^2 \Delta t + \left[\left(r_t - \frac{\sigma_t^2}{2}\right) \Delta t\right]^2. \end{aligned} \tag{2.18}$$

This method does not guarantee that the discounted stock price is a $Q$ martingale. The idea behind this method is that, for short time intervals, the behaviour of the observed logreturns may be well-approximated by a diffusion. The moment constraints in Equation 2.18 are derived using a series expansion of $e^{R_{t+1}}$ and ignoring

higher order terms. If the higher order terms can be ignored, then the risk-neutral moments of the observed logreturns should be similar to those from a diffusion. This method is computationally more stable than REB-L1 and is consistent with the Black-Scholes model when the returns are lognormal. The differences between REB-L1 and REB-L2 are examined in Section 2.3.2 and are shown to be insignificant for daily time intervals.

## 2.3 Mean Logreturn vs Mean Stock Price

In this section, we discuss the differences in the MRED resulting from moment constraints imposed on the (discounted) stock price versus the MRED obtained when the moment constraints are prescribed on the logreturns. We first discuss the typical assumption that the stock price at some future time $T$ is lognormally distributed. Another example investigates the REB models of the previous sections.

### 2.3.1 Lognormal Distribution

Let us assume that under the $P$ measure, the stock price at time $T$ is lognormally distributed. That is, assume that $S_T = S_0 e^{X_T}$ where $X_T \sim N(\mu T, \sigma^2 T)$, for some mean $\mu T$ and variance $\sigma^2 T$. Let $f^P_{S_T}$ and $f^P_{X_T}$ denote the lognormal density function of $S_T$ and the normal density function of $X_T$ under the $P$ measure respectively. Furthermore, assume that $r$ is the constant risk-free rate of return (continuously compounded).

The goal is to compute an asset pricing measure. The theory of financial economics asserts that under a risk-neutral measure the discounted asset price is a

martingale (a consequence of the no-arbitrage assumption). Two approaches using minimum relative entropy will be used to construct pricing measures. The first approach works directly with the lognormal density $f^P_{S_T}$ and moment constraints are imposed on the stock price. The second approach assumes that under $Q$, the logreturn is normally distributed and the martingale constraint dictates the required mean of the logreturn.

### Mean Stock Price

Relative entropy is used to select the density for $S_T$ that is closest to $f^P_{S_T}$ and satisfies the martingale constraint. Specifically, we seek the solution to

$$\begin{aligned} \min_{\tilde{f}} H(\tilde{f}|f^P_{S_T}) &= \min_{\tilde{f}} \left\{ \int_\Omega \tilde{f}(s) \ln \left( \frac{\tilde{f}(s)}{f^P_{S_T}(s)} \right) ds \right\} \qquad (2.19) \\ \text{subject to} \quad \int_\Omega \tilde{f}(s) ds &= 1, \\ E_{\tilde{f}}[S_T] &= S_0 e^{rT}, \\ \tilde{f} &\ll f^P_{S_T}. \end{aligned}$$

By Theorem 1.4 the solution to this problem is the MRED with density

$$\begin{aligned} f^{Q_1}_{S_T}(s) &= f^P_{S_T}(s) \exp\{\lambda s - \psi(\lambda)\} \qquad (2.20) \\ &= \frac{1}{s\sqrt{2\pi}\sigma\sqrt{T}} \exp\left\{ \frac{(\ln s - \mu T)^2}{2\sigma^2 T} + \lambda s - \psi(\lambda) \right\} \end{aligned}$$

where $\psi'(\lambda) = S_0 e^{rT}$. Clearly, $f^{Q_1}_{S_T}$ is not a lognormal density function. Also notice the appearance of the parameter $\mu$ in the density function. Both of these

observations are counter to the Black-Scholes model even though we started with a lognormal density for $S_T$ under the $P$ measure. The disappearance of $\mu$ in the Black-Scholes model is due to the argument of continuously rebalancing the hedging portfolio. This argument effectively keeps the stock price distribution lognormal under $Q$ and adjusts the mean logreturn to ensure that the martingale constraint is satisfied. However, when hedging is not continuously performed, the martingale constraint alone is not sufficient to ensure that the stock price distribution remains lognormal under $Q$, hence the appearance of $\mu$ in the risk-neutral density function.

### Mean Logreturn

Notice that, in the above argument we *do not* assume that, under $Q$, the stock price at time $T$ is lognormally distributed. If we make the assumption that the logreturn $X_T \sim N(\tilde{\mu}T, \tilde{\sigma}^2 T)$ in the risk-neutral world, then the martingale constraint on the discounted stock price implies that

$$\begin{aligned} & E^Q\left[e^{-rT}S_T\right] &=& S_0 \\ \Leftrightarrow\ & E^Q\left[e^{X_T - rT}\right] &=& 1 \\ \Leftrightarrow\ & e^{\tilde{\mu}T - rT + \frac{\tilde{\sigma}^2 T}{2}} &=& 1 \\ \Leftrightarrow\ & \tilde{\mu} &=& r - \tfrac{\tilde{\sigma}^2}{2}. \end{aligned} \tag{2.21}$$

Therefore, under this assumption the risk-neutral density of the logreturn, $f_{X_T}^{Q_2}$ is normal with mean $(r - \frac{\tilde{\sigma}^2}{2})T$ and variance $\tilde{\sigma}^2 T$. Gerber and Shiu (1994) use this approach of assuming that the logreturns have the same distributional form in the risk-neutral and physical worlds in their derivation of option pricing models using

the Esscher transform.

The risk-neutral distribution, $f_{X_T}^{Q_2}$ also solves the relative entropy optimisation problem

$$\begin{aligned} \min_{\tilde{f}} H(\tilde{f}|f_{X_T}^P) &= \min_{\tilde{f}} \left\{ \int_\Omega \tilde{f}(x) \ln \left( \frac{\tilde{f}(x)}{f_{X_T}^P(x)} \right) dx \right\} \qquad (2.22) \\ \text{subject to} \quad \int_\Omega \tilde{f}(x)dx &= 1, \\ E_{\tilde{f}}[X_T] &= \left( r - \frac{\sigma^2}{2} \right) T, \\ \tilde{f} &\ll f_{X_T}^P. \end{aligned}$$

Note that the variance of $X_T$ is unchanged giving $\tilde{\sigma}^2 = \sigma^2$. The change of variable $S_T = S_0 e^{X_T}$ gives the risk-neutral density for the stock price $f_{S_T}^{Q_2}$, which is lognormal. This method of constraining the mean logreturn and then making a change of variable agrees with the Black-Scholes model. Notice the absence of the parameter $\mu$ in the resulting risk-neutral distribution. In his PhD thesis, Gulko uses this method to come up with a generalised Black-Scholes pricing formula using entropy optimisation principles. However, it should be noted that this method assumes knowledge of the distributional form under the risk-neutral measure (Gulko (1998)).

To compare $f_{S_T}^{Q_1}$ with $f_{S_T}^{Q_2}$, we compute these densities assuming $S_0 = 1336.4, \mu = 0.15, \sigma = 0.22$ and $r = 0.04$ at a number of different maturities $T$. Figure 2.1 shows that for a time horizon of 1 day, these densities are virtually indistinguishable. Increasing the time horizon, we see that slight differences become apparent as $T$ increases (see Figures 2.2 and 2.3). We conclude that for short time horizons it does

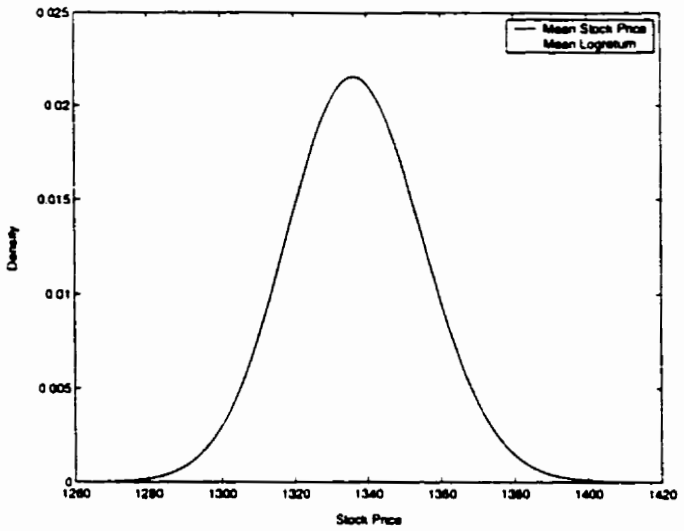


Figure 2.1: Comparison of risk-neutral densities for a 1 day horizon.

not matter which transformation to a pricing measure one uses, as they produce essentially the same distribution.

### 2.3.2 REB

Here, we examine the differences of the REB models with particular interest on the comparison between REB-L1 and REB-L2. To investigate this we use the S&P500 data described in Section 2.7.1. The risk-free rate is assumed to be $r = 0.05$ and each of the models is calibrated (see Sections 2.5 and 2.7.2) to the 10-day call options (in order for the volatility of each of the models to be *correct*). Figure 2.4 shows the cdf of $S_{10}$ based on 50,000 simulations for each of the REB models. Figure 2.5 gives the corresponding picture for the 29-day options. The cdf's from REB-L1 and REB-L2 are indistinguishable at each of the maturities. As in the previous section,

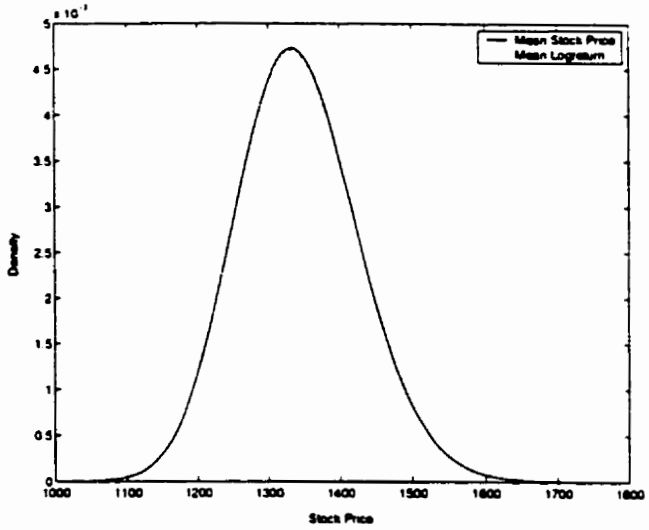


Figure 2.2: Comparison of risk-neutral densities for a 1 month horizon.

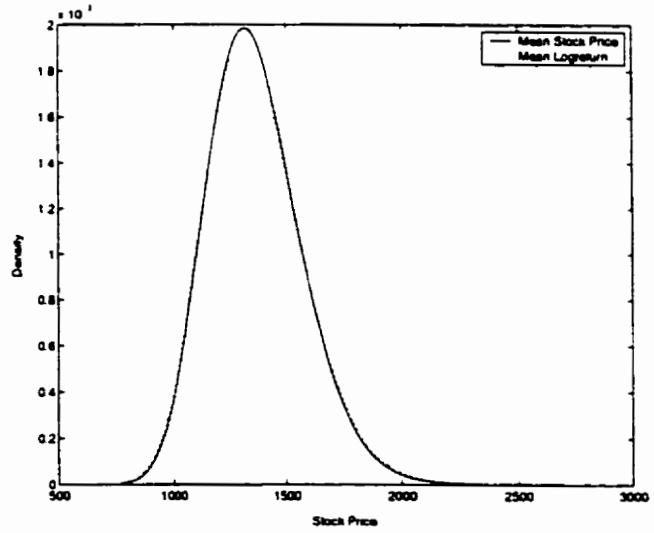


Figure 2.3: Comparison of risk-neutral densities for a 6 month horizon.

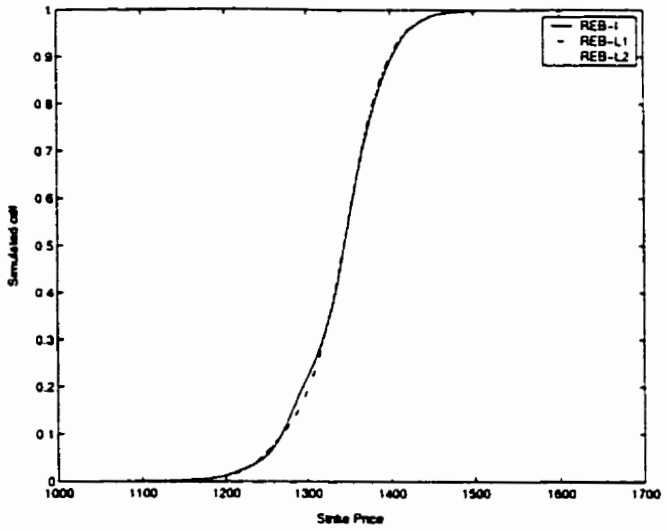


Figure 2.4: Comparison of REB risk-neutral cdfs for a 10 day horizon.

this shows that for short time increments it does not matter which transformation to a pricing measure one uses. REB-I has a slightly different cdf at the 10-day maturity than those produced with the other bootstrap models. This difference disappears at the longer maturity due to the effect of the CLT.

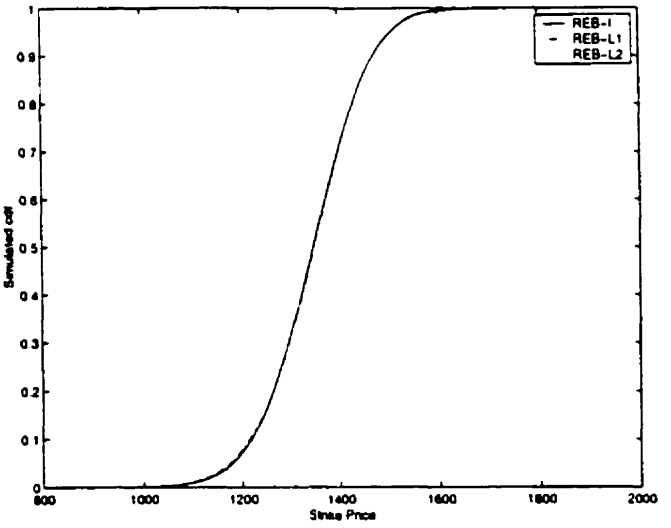


Figure 2.5: Comparison of REB risk-neutral cdfs for a 29 day horizon.

## 2.4 Stutzer's Model

Stutzer (1996) proposed a nonparametric option pricing model based on the price history of the underlying asset. In a certain sense, to be made clear in what follows, the REB models are generalisations of this approach. Although this model can be cast in the Markov market model, it is simpler not to invoke its heavy notational machinery.

As with the bootstrap models, we start with a history of daily closing prices, $s_0, \ldots, s_N$, and suppose we wish to price vanilla call options that mature in $T$ days. In this model, the first step in building the $P$ measure is to compute the (rolling) $T$-day gross returns from the price history. The observed $T$-day gross returns are

$$r(n) = \frac{s_{n+T}}{s_n}, \qquad \text{for} \quad n = 0, \ldots, N - T. \tag{2.23}$$

The set of unique observed $T$-day gross returns $\mathcal{R}$ is constructed and corresponding to each $r \in \mathcal{R}$ is the observed relative frequency, $w_r$, of that gross return. That is,

$$w_r = \frac{\sum_{n=0}^{N-T} I(r(n) = r)}{N - T}, \quad \text{for all} \quad r \in \mathcal{R}. \tag{2.24}$$

Suppose $S_0$ is the current stock price, then the set of possible stock prices at time $T$ is constructed by

$$S_T(r) = S_0 r, \qquad \text{for all} \quad r \in \mathcal{R}. \tag{2.25}$$

This determines the support of the $P$ measure and each of these outcomes is assigned probability equal to the observed relative frequency of the unique gross return. That is,

$$P(S_T = S_T(r)|S_0) = w_r. \qquad \text{for all} \quad r \in \mathcal{R}. \tag{2.26}$$

As with the REB models, relative entropy is minimised to transform this $P$ measure to a pricing measure, $Q$, subject to some prescribed moment constraints (e.g., the martingale constraint on the discounted stock price). As usual, prices for vanilla call options expiring at time $T$ are then computed as the expected discounted value of the payoff, with the expectation taken under the $Q$ measure.

It should now be clear that the REB models are dynamic versions of Stutzer's model. That is, the REB models construct a risk-neutral process from time 0 to time $T$, rather than a risk-neutral distribution at time $T$. Potential advantages of Stutzer's model over the REB models include the capturing of the $T$-day de-

pendence of the stock price by the $P$ measure, exact model option prices (i.e., no simulation error), and faster computations. However, the REB models produce a much richer set of possible prices at time $T$ and, since price paths are generated, REB models can be used to price European derivatives with arbitrary payoffs. Furthermore, REB models are able to price options at more than one maturity while Stutzer's model is valid only at a single maturity. In Section 2.7, REB models are shown to be superior to Stutzer's model when pricing S&P500 options at a single maturity.

## 2.5 Calibration

An important feature of any derivatives pricing model is the ability to calibrate the parameters of the model to accurately reflect the current market prices of traded derivatives. For example, the Black-Scholes call option pricing formula is a function, $BS(S_0, T, r, K, \sigma)$, of the current stock price $S_0$, the time until the option expires $T$, the risk-free rate $r$, the strike price $K$ and the volatility $\sigma$. With the exception of $\sigma$, all of these parameters are known directly from the market ($S_0$ and $r$) or from the specification of the option ($T$ and $K$). Suppose the current market price of this option is $C$. Then, the value of $\sigma$ that gives $C$ as the model price can be computed by numerically inverting the function $BS$. That is,

$$BS(S_0, T, r, K, \sigma_{imp}) = C, \tag{2.27}$$

and $\sigma_{imp}$ is called the Black-Scholes implied volatility.

Now, consider the more general framework in which there are $J$ call options at a single maturity with market prices $P_j$ corresponding to strike prices $K_j$ for $j = 1, \ldots, J$. We wish to price these options using some general pricing model that depends on a (possibly vector-valued) calibrating parameter $\theta$. Let $P_j(\theta)$ and $D_j(\theta) = P_j(\theta) - P_j$ be the model price at $\theta$ and the pricing error at $\theta$ for option $j$ respectively, for $j = 1, \ldots, J$. Also, let $P(\theta)$ and $D(\theta)$ be the corresponding $J \times 1$ column vectors.

Here, $\theta$ is a parameter of the pricing model whose value will be determined by calibrating the model to the observed market prices. In this section, we consider $\theta$ to be a positive scalar parameter that governs the second moment (i.e., volatility) of the logreturns or stock prices. Typically, when $\theta$ is a scalar, no single value exactly reproduces the market prices of all options at one maturity. However, a calibrating value of $\theta$ is determined by minimising, for example, the sum of squared pricing errors (SSPE) as a function of $\theta$ (other loss functions can also be used). Specifically,

$$\theta_{cal} = \arg\min_{\theta>0} D(\theta)' D(\theta) = \arg\min_{\theta>0} \sum_{j=1}^{J} [D_j(\theta)]^2, \tag{2.28}$$

where $(')$ denotes transpose.

For pricing models having a closed-form (e.g., Black-Scholes) the pricing error $D(\theta)$ and its gradient $\nabla D(\theta)$ are precisely known. Thus, the above optimisation problem is easily solved using a Newton-Raphson technique and we find that $\theta_{cal}$

satisfies

$$D(\theta)'\nabla D(\theta) = \sum_{j=1}^{J} D_j(\theta)\frac{\partial P_j(\theta)}{\partial \theta} = 0. \tag{2.29}$$

However, the calibration of simulated pricing models (e.g., REB models) is not as straightforward. For these models, the pricing error $D(\theta)$ is not known precisely and must be estimated via simulation. Of course, this estimate will contain simulation error. Furthermore, since the model is not in closed-form, analytic derivatives are unavailable. The lack of a gradient rules out many root finding or optimisation packages, while the simulation noise in the estimate of the pricing error presents another problem. These issues motivated the development of a method for calibrating simulated pricing models presented in Chapter 3. This algorithm is quite general as it can handle a vector of calibrating parameters as well as options covering a range of strike prices and maturities. Please refer to Chapter 3 for more details.

## 2.6 Simulated Pricing Models

### 2.6.1 Bootstrap Valuation of European Options

Suppose we wish to value an abritrary European option paying $g_T$ at time $T$, where $g_T$ is necessarily $\mathcal{F}_T$-measurable. For example, a plain vanilla call option with strike $K$ expiring at time $T$ has the payoff function $g_T = (S_T - K)^+ = \max(S_T - K, 0)$. Another example is a path-dependent option paying $g_T = (S_T - \bar{S})^+$, where $\bar{S}$ is

the average closing value of the stock from $t = 0$ to $t = T$.

In the language of Section 1.6, $g_T$ is the functional of the distribution of $S_T$ that we want to estimate (for ease of notation, we will suppress the distributional dependence of $g_T$). Starting with the current price $S_0$, one of the REB models is used to generate a bootstrap stock price path $(S_0, S^*_{1,1}, \ldots, S^*_{T,1})$. The option payoff based on this path is $\hat{g}^*_{T,1}$ (note that $\hat{g}^*_{T,1}$ is $\mathcal{F}^*_{T,1}$-measurable, where $\mathcal{F}^*_{T,1} = \sigma(S_0, S^*_{1,1}, \ldots, S^*_{T,1})$).

Suppose we generate $B$ bootstrap sample paths, then the bootstrap estimate of $g_T$ is given by

$$\hat{g}^*_T = \frac{\sum_{b=1}^{B} \hat{g}^*_{T,b}}{B}, \tag{2.30}$$

where $\hat{g}^*_{T,b}$ is the estimate of $g_T$ from the $b$th bootstrap sample path. The bootstrap estimate of the current price of the option is just $\hat{g}^*_T$ discounted by the risk-free rate. Obviously the bootstrap is equivalent to Monte Carlo estimation where the random variables are drawn from an empirical distribution function, rather than from some parametric family.

### 2.6.2 Hedging Parameters

Hedging parameters for simulated pricing models can easily be estimated from the simulation. For example, to calculate the sensitivity of an option to a change in the underlying stock price (i.e., the option's $\Delta$) first assume the initial stock price is $S_0$ and estimate the option price $\hat{P}(S_0)$ by simulation. Then, using the same random number stream estimate the option price assuming the initial stock price is $S_0 + h$

for some small positive number $h$. Call this estimate $\hat{P}(S_0 + h)$. The option's $\Delta$ is then estimated by

$$\hat{\Delta} = \frac{\hat{P}(S_0 + h) - \hat{P}(S_0)}{h}. \tag{2.31}$$

Other hedging parameters can be estimated in a similar manner (Hull (1997)). This is the method used to calculate the $\hat{\Delta}$'s for REB-I in Section 2.7.

For REB-L1 and REB-L2, the stock price at maturity can be written as

$$S_T = S_0 X_T, \tag{2.32}$$

for some random variable $X_T$ whose distribution does not depend on $S_0$. Assuming that the interchange of limits is allowed, the $\Delta$ of a call option with strike $K$ is estimated by

$$\hat{\Delta} = \hat{E}\left[X_T e^{-rT} I\left(X_T \geq \frac{K}{S_0}\right)\right], \tag{2.33}$$

where $\hat{E}[\cdot]$ denotes expectation with respect to the simulated risk-neutral measure. This method is used to calculate the $\hat{\Delta}$'s for REB-L1 and REB-L2 in Section 2.7.

Another approach for estimating hedging parameters for simulated pricing models has recently been proposed by Avellaneda et al (1999) and Avellaneda and Gamba (2000). It may be interesting to use these methods with the bootstrap models.

## 2.7 Example: S&P500 Index Options

### 2.7.1 Background Information

The REB models were used to price European call options on the S&P500 index. The S&P500 was chosen due to the availability of historical data on the index and the current market prices of a wide range of index options (i.e., options with different strikes and maturities). Although the S&P500 is not an ideal choice from the perspective of modelling the price discreteness, we still gain the advantage of valuing options without any parametric distributional assumptions on the evolution of the index.

The historical data used in our analysis is the daily closing level of the S&P500 index from March 15, 1991 to May 7, 1999. Note that this time period is completely arbitrary and issues such as the time period studied and the length of the price history are not investigated here. The market data used are the bid/ask spreads of traded S&P500 call options at 3:30pm on May 10, 1999 and the closing value of the index on May 10 is 1336.43. In what follows, we take the midpoint of the bid/ask spread as the market price of each option.

Options are priced at a single maturity and we assume that the risk-free rate and volatility are constant, that is, $r_t = r$ and $\sigma_t = \sigma$ for all $t$. This allows for valid comparisons of the REB, Black-Scholes and Stutzer pricing models. The risk-free rate is taken from the market as the rate earned on the Treasury bill that expires nearest to the option expiration date ($r = 0.05$). Here, we ignore the dividend stream on the index as it will have little effect on the short-term options priced in

| Calibrating Volatility | | | | |
|---|---|---|---|---|
| Model | 10-day Maturity | | 29-day Maturity | |
| Black-Scholes | 0.1944 | | 0.2213 | |
| Stutzer | 0.2060 | | 0.2350 | |
| | $\Delta t = 1$ | $\Delta t = 2$ | $\Delta t = 1$ | $\Delta t = 5$ |
| REB-I | 0.2011 | 0.1980 | 0.2193 | 0.2129 |
| REB-L1 | 0.1970 | 0.2003 | 0.2220 | 0.2169 |
| REB-L2 | 0.2020 | 0.2023 | 0.2273 | 0.2272 |

Table 2.1: Calibrating volatilities for S&P500 call options.

this example.

For both maturities studied here, the bootstrap models use 1-day time intervals. To try and capture some dependence and elude the effects of the CLT by summing fewer terms, we also use 2-day time intervals for the 10-day options and 5-day (weekly) time intervals for the 29-day options. We price the 29-day options as if they are 30-day options, ignoring the effect of the extra day.

### 2.7.2 Results

#### Calibration

For each maturity and for each pricing model, a calibrating volatility is computed. The values for the Black-Scholes and Stutzer models are computed exactly via Newton-Raphson while the calibration algorithm of Chapter 3 is used for the REB models. The calibrating volatilities are given in Table 2.1.

### Pricing Errors of Traded Options

A convenient way to assess pricing models is to plot the pricing error (the model price minus the market price) against the strike price. Figure 2.6 is a plot of the pricing error of the models against the strike price for the 10-day call options using $\Delta t = 1$ in the REB models. Figure 2.7 is the pricing error plot using $\Delta t = 2$. Figures 2.8 and 2.9 are the corresponding plots for the 29-day call options with $\Delta t = 1$ and $\Delta t = 5$ respectively. At both maturities, the REB model prices are computed using 50,000 simulations.

From these plots, it is easily seen that the REB models do a better job at reproducing the market prices of the options than the Black-Scholes and Stutzer models. The REB models using time intervals of 2 and 5 days outperform the REB models with $\Delta t = 1$. Figure 2.8 shows that the REB ($\Delta t = 1$) and Black-Scholes models perform well at-the-money but produce significant pricing errors for in-the-money and out-of-the-money options. The Stutzer model performs poorly as it underprices at-the-money options and overprices in-the-money and out-of-the-money options. Due to this poor performance, we decided not to estimate hedging parameters for this model. Similar pricing error patterns to those observed for the 29-day options emerge when pricing options at longer maturities using 1-day time intervals.

REB models seem to have an advantage over Black-Scholes at shorter maturities. However, as the maturity increases this advantage tends to disappear. At longer maturities more price movements are added together in the bootstrap models. The more terms added together, the stronger the effect of the CLT. Therefore, at longer

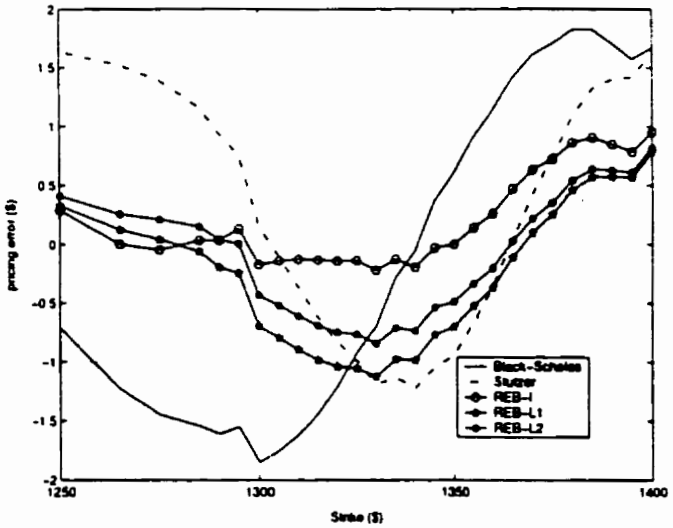


Figure 2.6: Pricing Errors for 10-day S&P500 call options ($\Delta t = 1$).

maturities the CLT governs the distribution of $S_T$, rendering the REB models similar to Black-Scholes.

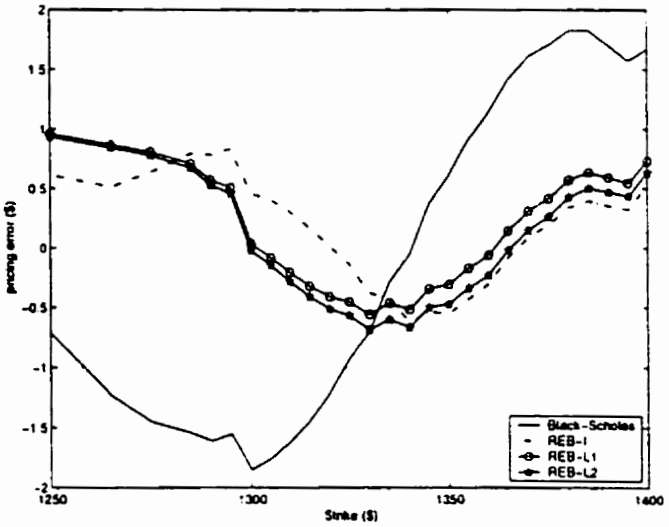


Figure 2.7: Pricing Errors for 10-day S&P500 call options ($\Delta t = 2$).

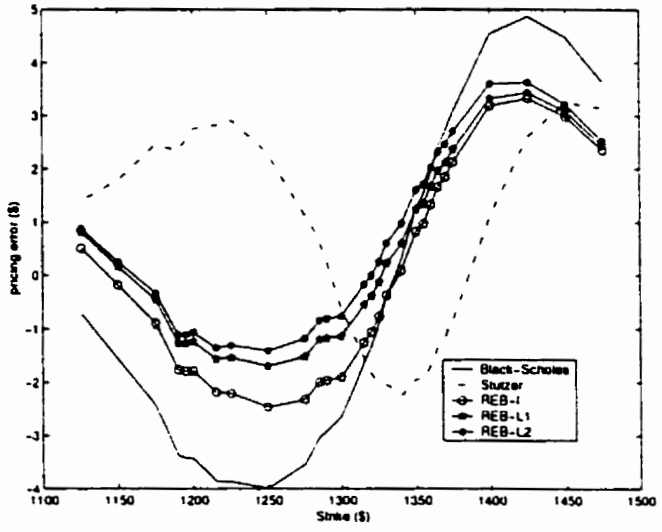


Figure 2.8: Pricing Errors for 29-day S&P500 call options ($\Delta t = 1$).

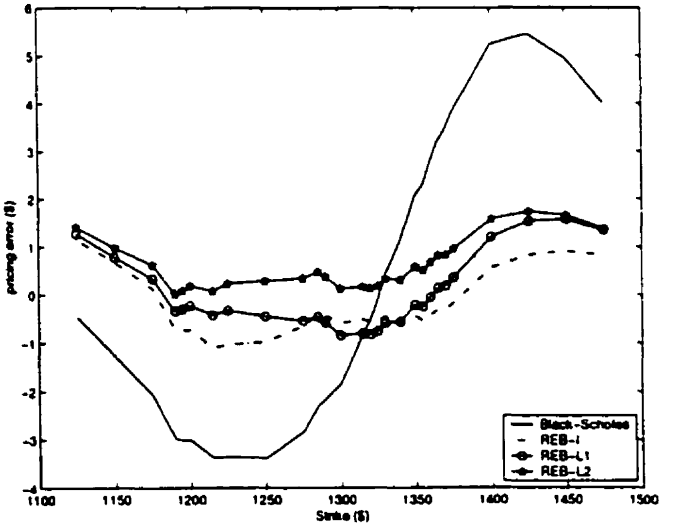


Figure 2.9: Pricing Errors for 29-day S&P500 call options ($\Delta t = 5$).

### Hedging Parameters

Hedging parameters were estimated using the methods outlined in Section 2.6.2. Figures 2.10 and 2.11 plot the $\hat{\Delta}$'s for the 10-day options corresponding to $\Delta t = 1$ and $\Delta t = 2$ respectively. Figures 2.12 and 2.13 are the corresponding plots for the 29-day options with $\Delta t = 1$ and $\Delta t = 5$ respectively. These estimates are based on the same simulations as those used to price the options. The estimates for REB-I were produced using $h = 0.01S_0$, an arbitrarily chosen perturbation of 1% of the initial index level. There are some differences between the $\hat{\Delta}$'s for the REB models and those from the Black-Scholes model. The differences are most apparent for the 10-day options and the 29-day options with $\Delta t = 5$. Since the REB models are more consistent with the observed option prices in these cases, REB-estimated hedging parameters may have an advantage over their Black-Scholes counterparts. At longer

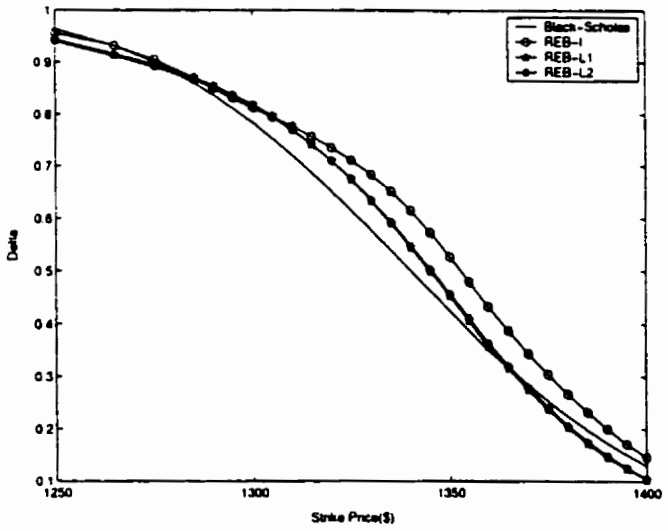


Figure 2.10: $\hat{\Delta}$'s for 10-day S&P500 call options ($\Delta t = 1$).

maturities or using short time increments, this advantage tends to disappear due to the effect of the CLT.

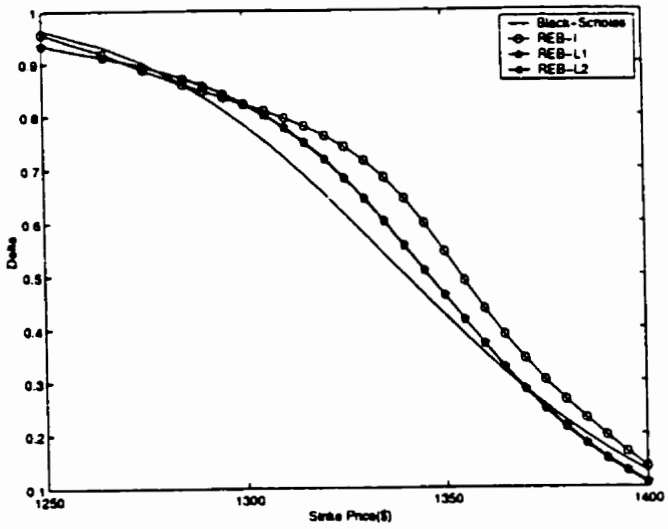


Figure 2.11: $\hat{\Delta}$'s for 10-day S&P500 call options ($\Delta t = 2$).

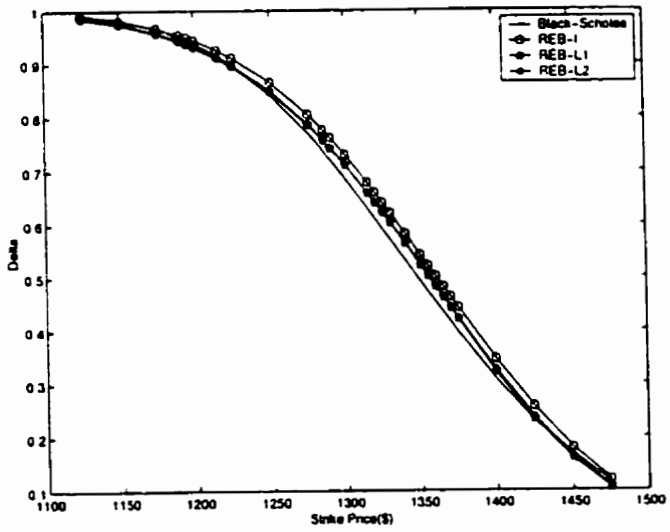


Figure 2.12: $\hat{\Delta}$'s for 29-day S&P500 call options ($\Delta t = 1$).

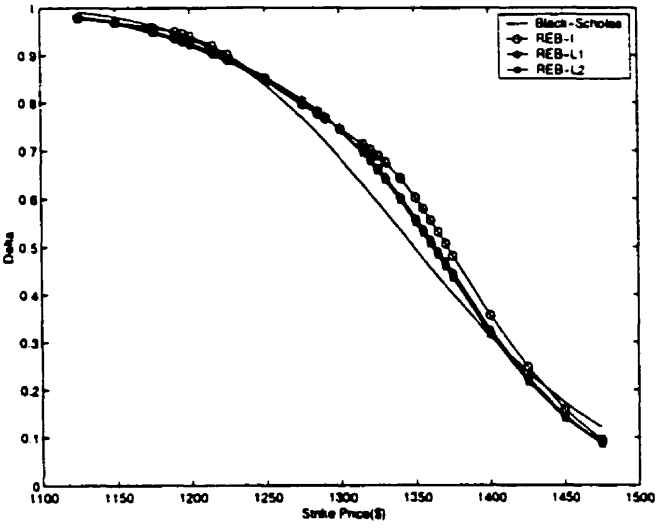


Figure 2.13: $\hat{\Delta}$'s for 29-day S&P500 call options ($\Delta t = 5$).

### 2.7.3 Discussion

The REB models outperform the Black-Scholes and Stutzer models, particularly at a short maturity. At the longer maturities studied to date, the REB with $\Delta t = 1$ and Black-Scholes models perform well at-the-money, but have significant pricing errors otherwise. Using longer time increments in the REB models yields an improvement in pricing performance. On the other hand, Stutzer's model performs poorly at-the-money and, since most of the market activity involves at-the-money options, we conclude that this model is not appropriate for the options studied here. The dynamic REB models are clearly more consistent with the options considered here than the static nonparametric model of Stutzer.

Adjusting the REB models to time increments of $\Delta t = 2$ and $\Delta t = 5$ has two desirable features. First, possible autocorrelations in the daily price movements are

captured and, since fewer terms are summed, the adjustment also helps to stave off the effect of the CLT.

The pricing error pattern exhibited by the REB ($\Delta t = 1$) and Black-Scholes models (under pricing in-the-money options and overpricing out-of-the-money options) indicates that a distribution that is more skewed and heavier-tailed is required to accurately reproduce the market prices of options. This pattern is consistent with that observed by Dumas, Fleming and Whaley (1998). Heston (1993) shows how this type of pattern can be reproduced using a stochastic volatility model that is correlated with the asset price movements.

### Benefits of REB Models

REB models are completely nonparametric models that rely on the historical distribution of the asset, rather than imposing some inappropriate parametric model on the stock price dynamics. Desirable features of asset price movements such as fat tails and skewness are automatically incorporated into the model, provided that these features have been observed in the past. The REB-I model provides a way to model asset price movements on the original scale and in such a manner as to preserve the discrete price jumps.

Compared with Stutzer's model, the REB models produce a much richer distribution at a given maturity. We have shown that this feature gives the REB models a pricing advantage over the Stutzer's model. Since a risk-neutral process rather than a distribution is generated, there are other obvious benefits of the REB compared to Stutzer's model. For example, one can value European securities with path-dependent payoff functions (e.g., barrier options). Furthermore,

both stochastic interest rate and volatility models can be incorporated into the REB models with minimal difficulty. We need only adjust the required conditional mean and variance at each time point to current estimates of the risk-free rate and volatility. A (stochastic) dividend model can also be similarly incorporated into the REB models. Clearly, the extension of the static model proposed by Stutzer to the dynamic REB models has many potential benefits.

### Drawbacks of REB Models

REB models tacitly assume that the distribution of future movements in the asset price is completely supported by the empirical distribution of the observed movements (increments for REB-I and logreturns for REB-L1 and REB-L2). In other words, when simulating bootstrap price paths, only changes that have been observed in the past are possible in the simulated paths. One can argue that, provided there is a sufficient price history, this problem is not too severe and that the gains from the bootstrap approach outweigh this potential problem.

As previously mentioned, the assumption of a random walk on the observed increments (REB-I) may also be invalid. Typically, large changes in the price of the asset occur more frequently when the asset price is high than when it is low. This can be particularly problematic if the majority of the observed increments occurred at prices much different than the current value of the asset. Furthermore, with the additive random walk assumption there is positive probability that the asset price can become negative. This problem has not been encountered in any examples studied to date.

Another drawback of REB models is the significant amount of computational

time required to estimate a price. Attempts were made to improve the efficiency of the REB models through commonly-known variance reduction techniques. In fact, antithetic and control variates, importance sampling and best linear unbiased estimators (BLUE) were all tried with limited success. Antithetic variables were the most successful at reducing variance, although the gains in efficiency were miniscule. Martingale variance reduction techniques (Clewlow and Carverhill (1995)) were also tried, but again only a minor gain in efficiency was realised. Because of the intricate nature of the REB simulations (REB-I in particular), the best variance reduction technique may be a function of many things such as the strike price, time to maturity, volatility and risk-free rate. Thus, the best method will probably vary from case to case.

## 2.8 Hull-White Simulation Study

The purpose of this section is to investigate one of the possible causes for the pricing error pattern for the 29-day call options in Figure 2.8. As stated earlier, this pattern has also been observed in other studies on S&P500 options. Furthermore, Heston (1993) shows how this pattern may be induced by using a stochastic volatility model that is negatively correlated with the stock price.

To see if stochastic volatility is a possible explanation for the observed REB pricing errors, we performed a simulation study using the Hull-White stochastic volatility stock price model (Hull and White (1987)). This model is explained in more detail in Section 3.3.2. The same parameter values as those used in Section 3.3.2 for the zero-drift variance process are used here. The parameter values are

$S_0 = 1336.43$, $\Delta t = \frac{1}{252}$, $r = 0.04$, $\sigma_0^2 = 0.1326$, $\rho = -0.5942$ and $\xi = 0.7934$. A 5 year history of daily closing values is generated with these parameter values and the 20-day option prices given in Section 3.3.2 are used as the market option prices. The simulated 5 year history of closing prices is used to define the $P$ measure for the REB models.

We proceed as in Section 2.7.2, assuming a constant risk-free rate of 0.04 and a constant volatility. The REB models are then calibrated to the simulated option prices via the method in Chapter 3 and the pricing error is plotted against strike price in Figure 2.14. The pricing error pattern observed here is consistent with that obtained for the 29-day S&P500 options. This suggests that a model with changing volatility may have been more appropriate when pricing this set of options. In fact, in Section 3.4.3 it is shown that the Hull-White stochastic volatility model is appropriate for the 29-day S&P500 options. This example also highlights the need to develop a bootstrap model for changing volatility (possibly by using observed squared returns), something that is left for future work.

## 2.9 Assumption of Independence

Results of this chapter call into question the assumption of the independence of daily stock price movements (either increments or logreturns) in the risk-neutral world. Indeed, the simulation study in the previous section suggests that some of the dependence may be captured using a model of changing volatility. However, in this section we continue to assume that volatility is constant. We investigate whether any model that assumes the risk-neutral independence of the logreturns

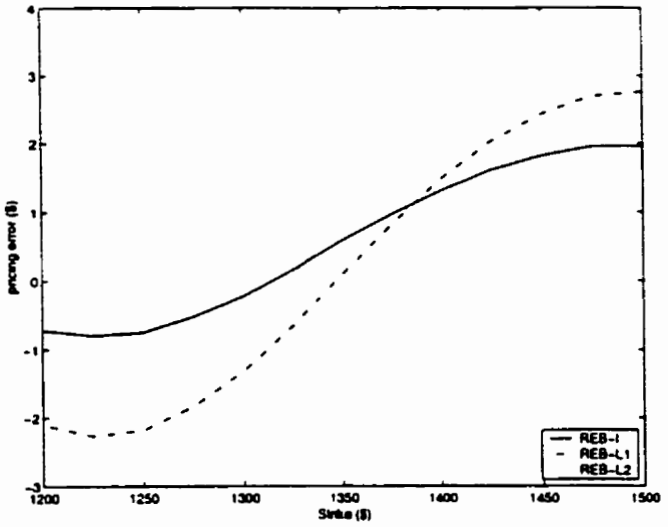


Figure 2.14: Pricing errors for 20-day Hull-White simulated call options.

can produce a risk-neutral process for the underlying stock price that is consistent with option prices at a number of different maturities.

### 2.9.1 Implied Risk-neutral Distributions

In this section, we calculate implied risk-neutral distributions of the S&P500 index from the market prices of call options. This is done at a number of different maturities. As mentioned in Section 1.3.2, many methods can be used to solve this problem. In keeping with the spirit of this thesis, we shall use relative entropy to determine the implied risk-neutral distributions. The implied skewness of these distributions is calculated, revealing some interesting implications.

**Methodology**

Suppose at a single maturity, $T$, we have the market prices of $J$ call options, $C_1, \ldots, C_J$, corresponding to strike prices $K_1, \ldots, K_J$. Let $S_0$ and $r$ be the current stock price and the risk-free rate respectively. We use the model

$$S_T = S_0 e^{X_T} \tag{2.34}$$

to describe the $S_T$ in terms of its logreturn $X_T$ and take a lognormal distribution for $S_T$ (i.e., a normal distribution for $X_T$) as a candidate for the risk-neutral distribution. Obviously, this is the usual Black-Scholes setup. By calculating a calibrating volatility $\hat{\sigma}$ as in Section 2.7.2 for this set of options, the candidate risk-neutral distribution for $\ln(S_T)$ is

$$\ln(S_T) \sim N\left(\ln(S_0) + \left(r - \frac{\hat{\sigma}^2}{2}\right)T, \hat{\sigma}^2 T\right). \tag{2.35}$$

Call the corresponding density function $g$.

The implied risk-neutral distribution, $F$, is defined as the MRED with density

$f$ that solves

$$\begin{aligned} \min_{\tilde{f}} H(\tilde{f}|g) &= \min_{\tilde{f}} \left\{ \int_\Omega \tilde{f}(x) \ln \frac{\tilde{f}(x)}{g(x)} dx \right\} \qquad (2.36) \\ \text{subject to} \quad \int_\Omega \tilde{f}(x) dx &= 1, \\ E_{\tilde{f}} \left[ e^{-rT} S_0 e^{X_T} \right] &= S_0 e^{rT}, \\ E_{\tilde{f}} \left[ e^{-rT} (S_0 e^{X_T} - K_j)^+ \right] &= C_j, \quad \text{for} \quad 1 \le j \le J, \\ \tilde{f} &\ll g. \end{aligned}$$

By Theorem 1.4 the solution (if one exists) is given by

$$f(x) = g(x) \exp \left\{ \sum_{j=0}^{J} \lambda_j e^{-rT} (S_0 e^x - K_j)^+ - \psi(\lambda_0, \ldots, \lambda_J) \right\}, \qquad (2.37)$$

where $K_0 = 0$, $C_0 = S_0 e^{rT}$, and $\frac{\partial \psi}{\partial \lambda_j} = C_j$ for $j = 0, \ldots, J$.

### 2.9.2 Example: S&P500

We have the market prices (midpoint of the bid/ask spread) of both call and put options on the S&P500 at a number of different maturities (the maturities are 29, 50, 93, 281 and 407 days). Using the method described above, we compute an implied risk-neutral distribution at each of these maturities.

Put-call parity was used to determine the implied risk-free rate minus the dividend stream giving $r - \delta = 0.04$. This is the typical adjustment for continuous dividends and is used in place of the risk-free rate. As we are pricing long-term options, it was decided to be more careful than in the previous examples in deter-

| Maturity ($T$ in days) | $\hat{\gamma}_{\ln(S_T)}$ |
|---|---|
| 29 | -0.6659 |
| 50 | -0.6511 |
| 93 | -0.7177 |
| 281 | -0.8458 |
| 407 | -0.6989 |

Table 2.2: Skewness of the implied risk-neutral distributions.

mining the correct discounting factor.

At each maturity, only those call options with an implied Black-Scholes volatility are used. At some of the maturities, there are a large number of options and, in order to speed the computation, some options are arbitrarily excluded from the set of moment constraints. At each maturity, the implied distribution was used to price the entire set of options and we found that the excluded options were priced correctly.

Figures 2.15 to 2.19 are the implied risk-neutral distributions of $\ln(S_T)$ at each maturity. It is clear that all of these distributions have a significant negative skew, evidence of the crash-o-phobia effect. Table 2.2 gives the skewness of the implied risk-neutral distribution (of $\ln(S_T)$) at each maturity. Notice the persistence in the skewness across maturities. From this observation alone, it is obvious that the CLT is not governing these distributions. The implied kurtosis at each maturity was approximately 3. This is due to the fact that outside the range of the strike prices, the options do not provide information about the tails of the risk-neutral distribution. Since we started with a normal distribution, the implied distribution essentially has normal tails.

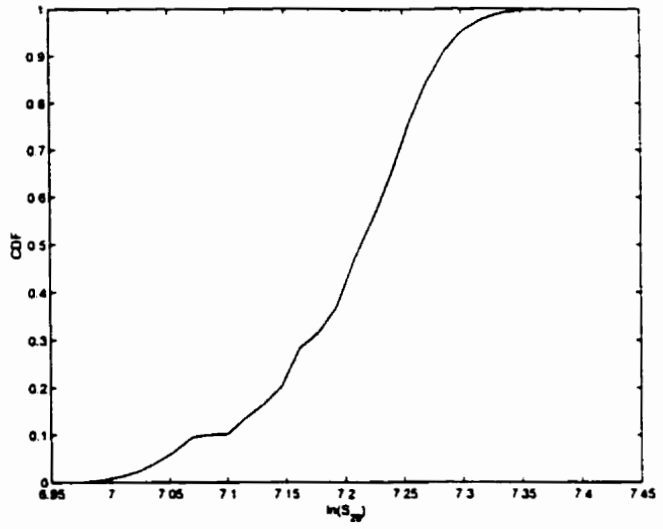


Figure 2.15: Risk-neutral distribution for $\ln(S_{29})$ implied from the 29-day S&P500 call options.

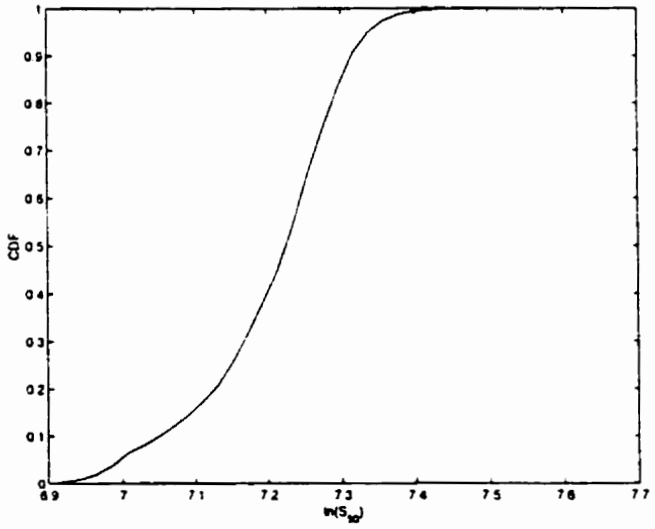


Figure 2.16: Risk-neutral distribution for $\ln(S_{50})$ implied from the 50-day S&P500 call options.

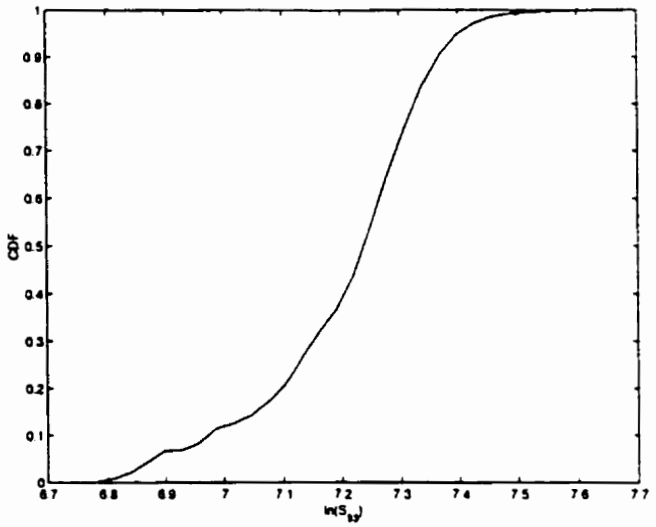


Figure 2.17: Risk-neutral distribution for $\ln(S_{93})$ implied from the 93-day S&P500 call options.

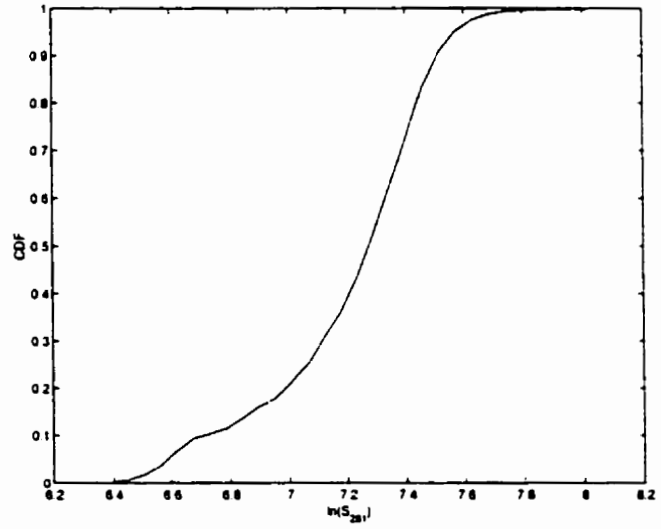


Figure 2.18: Risk-neutral distribution for $\ln(S_{281})$ implied from the 281-day S&P500 call options.

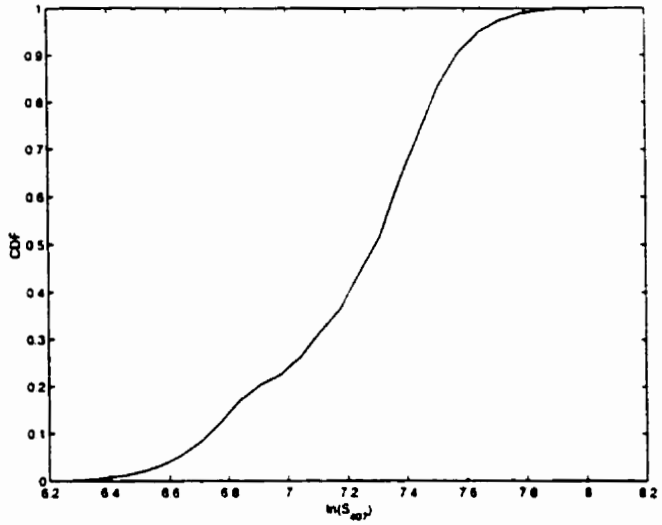


Figure 2.19: Risk-neutral distribution for $\ln(S_{407})$ implied from the 407-day S&P500 call options.

### 2.9.3 Risk-neutral Moments

Here, we use a discrete-time model for the risk-neutral evolution of the logarithm of the stock price. In particular time is discretised into intervals of equal length and that at time $T$ (i.e., after $T$ time intervals)

$$\ln(S_T) = \ln(S_0) + \sum_{t=1}^{T} Y_t, \tag{2.38}$$

where $S_0$ is the current stock price and $Y_t$, $t = 1, \ldots, T$ are i.i.d. random variables with some distribution $F_Y$. We assume that

$$E[Y] = \mu_Y \quad , \quad E[Y^2] = \sigma_Y^2 + \mu_Y^2, \tag{2.39}$$
$$E\left[\left(\frac{Y - \mu_Y}{\sigma_Y}\right)^3\right] = \gamma_Y \quad \text{and} \quad E\left[\left(\frac{Y - \mu_Y}{\sigma_Y}\right)^4\right] = \kappa_Y.$$

Here, the parameters $\gamma_Y$ and $\kappa_Y$ are the skewness and kurtosis of $F_Y$. Note that this model contains the (discrete-time) Black-Scholes model as a special case when the distribution $F_Y$ is normal.

Under the assumption that the logreturns are i.i.d., the skewness and kurtosis of $\ln(S_T)$ can be expressed in terms of the corresponding quantities of the logreturns. After some simple algebra, we obtain

$$\gamma_{\ln(S_T)} = E\left[\left(\frac{\ln(S_T) - \mu_Y T}{\sigma_Y \sqrt{T}}\right)^3\right] = \frac{\gamma_Y}{\sqrt{T}} \tag{2.40}$$

and

$$\kappa_{\ln(S_T)} = E\left[\left(\frac{\ln(S_T) - \mu_Y T}{\sigma_Y \sqrt{T}}\right)^4\right] = \frac{\kappa_Y}{T} + \frac{3(T-1)}{T}. \tag{2.41}$$

Now, since we are summing up $T$ i.i.d. random variables with finite variance, the CLT implies that

$$\ln(S_T) \sim AN\left(\ln(S_0) + \mu_Y T, \sigma_Y^2 T\right). \tag{2.42}$$

where $AN$ denotes asymptotically normal. This shows that even for moderate values of $T$ (e.g., 30 days), the skewness and kurtosis of $F_Y$ have little effect on the distribution of $\ln(S_T)$. Looking at Equations 2.40 and 2.41, we see that $\gamma_{\ln(S_T)} \to 0$ at a rate of $\frac{1}{\sqrt{T}}$ and $\kappa_{\ln(S_T)} \to 3$ at a rate of $\frac{1}{T}$. Of course, these limiting values are those of a normal distribution. This is the effect we saw when pricing the 29-day options using the REB models with $\Delta t = 1$.

Given a set of traded options at maturing at time $T$, suppose that one wants

to find a calibrating value not only for the volatility, but also for the skewness and kurtosis. Let $\hat{\gamma}_{\ln(S_T)}$ and $\hat{\kappa}_{\ln(S_T)}$ be these calibrating values. Rearranging Equations 2.40 and 2.41 we obtain the calibrating skewness and kurtosis of the logreturns, namely,

$$\hat{\gamma}_Y = \sqrt{T}\hat{\gamma}_{\ln(S_T)} \quad \text{and} \quad \hat{\kappa}_Y = \left(\hat{\kappa}_{\ln(S_T)} - 3\right) T + 3. \tag{2.43}$$

If $\hat{\gamma}_{\ln(S_T)}$ is much different from 0 or $\hat{\kappa}_{\ln(S_T)}$ is much different from 3, then $\hat{\gamma}_Y$ and $\hat{\kappa}_Y$ may impose unreasonably extreme (or unachievable) skewness and kurtosis on the distribution of $F_Y$, especially as the maturity date increases. This is the case for the S&P500 options with skewness hovering around -0.7 at each maturity.

Note that similar results can be derived when the i.i.d. assumption on the logreturns is relaxed. One need only apply a generalised CLT to show that $\ln(S_T)$ has a risk-neutral distribution that is asymptotically normal. For example, Theorem 35.12 in Billingsley (1995) is a martingale CLT that applies under some very mild distributional assumptions.

## 2.10 Conclusion

To summarise, the REB models are a dynamic version of Stutzer's nonparametric option pricing model. We have shown that the REB models outperform both the Black-Scholes and Stutzer models in pricing S&P500 options. At short maturities the bootstrap models are significantly better than Black-Scholes. However, at longer maturities this advantage disappears due to the effect of the CLT. The

persistent skewness of the implied risk-neutral distributions in Section 2.9.2 clearly show that any model consistent with the full set of options must have a feature that offsets the CLT.

In Chapter 3 we show that the Hull-White stochastic volatility model with correlation between the variance and stock price processes is consistent with traded S&P500 options at a single maturity. However, this model (with constant parameters) fails across maturities leading us to believe that the persistent negative skew likely comes from a risk-neutral jump component corresponding to the crash-o-phobia effect. In fact Andersen and Andreason (1999) show that the risk-neutral process in continuous time has a significant jump component.

# Chapter 3

# A Calibration Algorithm for Simulated Pricing Models

## 3.1 Introduction

Many pricing models require simulation in order to evaluate asset prices. The reason these models require simulation is that they are so complicated that no analytic solution or formula exists. Typically, these models rely on a vector of parameters that governs, for example, the dynamics of some underlying stochastic variables.

There is a great deal of econometric literature dealing with the estimation of these parameters under the physical measure $P$ (e.g., time series analysis and generalised method of moments (GMM) (Campbell et al (1997))). However, these estimates are typically not appropriate when pricing options. For example, the Black-Scholes implied volatility is typically higher than the volatility estimated from historical information. Carr et al (2000) also note that the parameter val-

ues for their process estimated under the $P$ measure are much different than the parameter values obtained from calibrating the model to option prices (i.e., estimating the parameters under a risk-neutral measure $Q$). Pricing models having an analytic form can be calibrated using well-known numerical techniques such as Newton's method, while the calibration of simulated models is a much more difficult task.

GMM-type methods have been applied to simulation-based estimators of asset pricing models, referred to as simulated moment methods (SMM) (Duffie and Singleton (1993)) or the method of indirect inference (Gouriéroux et al (1993)). These procedures use times series on the underlying asset and possibly related asset prices to glean information about the parameters of the data generating process. The methods that use time series option price information may be thought of as calibration procedures. The reader is referred to Gallant and Tauchen (1996), Christensen and Kiefer (2000), Gouriéroux et al (2000), Martin and Pagan (2000), McFadden (1989) and Pakes and Pollard (1989) for more information on simulation estimators, SMM and indirect inference. Recently Avellaneda et al (2001) have developed a method for calibrating simulated pricing models by reweighting the sample paths from the simulation to correctly price a set of benchmark assets. However, parameter estimates of the underlying data generating process are not directly produced by this method.

Pastorello et al (1999) use the indirect inference procedure to estimate the parameters of a stochastic volatility model for the underlying asset. In particular, they show that using at-the-money implied volatilities provide more accurate es-

timates of the parameters of the variance process than estimates obtained using information only on the underlying asset. However, they note the absence of a method for handling options with different moneyness, as well as the need to relax some of the restrictions of the structural model they employ. Specifically, they restrict themselves to the case where the stock returns and changes in volatility are uncorrelated.

The method presented here addresses both of these issues by incorporating options of different moneyness and allowing for correlation between stock returns and volatility. However, this method is a cross-sectional approach in the sense that no time series data on the options and the underlying are employed. Adapting this method to include time series on the underlying and the set of options is left for future work. The information set consists of the prices of European options, across a range of strikes and maturities, and the level of the underlying asset recorded at the same time.

## 3.2 Method of Calibration

### 3.2.1 The Setup

Consider a set of $J$ European options with market prices $P_j, j = 1, \ldots, J$. These options may have different maturities and denote the expiry date of the longest term option by $T$. Valuation of these options by simulation requires a model for the underlying stochastic variables, with a $(p \times 1)$ parameter vector $\theta \in \Theta$ governing the behaviour of these variables ($\Theta$ is the parameter space).

Let $P_j(\theta)$ be the model price for option $j = 1, \ldots, J$ at parameter value $\theta$. Also, let $D_j(\theta) = P_j(\theta) - P_j$ be the pricing error for option $j = 1, \ldots, J$ at parameter value $\theta$ and let $P(\theta)$ and $D(\theta)$ be the corresponding $(J \times 1)$ vectors. The objective is to calibrate this model to the market prices by finding the parameter $\hat{\theta}$ that minimises the sum of squared pricing errors (SSPE). In other words, $\hat{\theta}$ is the solution to

$$\hat{\theta} = \arg\min_{\theta \in \Theta} D(\theta)' D(\theta) = \arg\min_{\theta \in \Theta} \sum_{j=1}^{J} [D_j(\theta)]^2. \tag{3.1}$$

This is the familiar nonlinear least-squares problem studied extensively in statistics. In the analytic setting, this optimisation problem is solved by equating the derivative of the objective function to zero and solving for $\theta$. In other words, $\hat{\theta}$ solves

$$2D(\theta)' \nabla D(\theta) = 0. \tag{3.2}$$

Solution of this problem requires the functional form of both $D$ and $\nabla D$. For simulation-based models, this information is not available and it is proposed to approximate $D$ with a polynomial function of $\theta$. This approach of modelling the pricing error from which loss statistics can be predicted and optimised is consistent with Welch et al (1990) who use this approach in the area of quality improvement. The idea is that the individual responses are much simpler and hence easier to model accurately than loss functions.

The loss function $D(\theta)' D(\theta)$ can be generalised to $D(\theta)' \Omega D(\theta)$ where $\Omega$ is a suitable weighting matrix. Information such as moneyness, the option price, trading

volume, open interest and time-to-maturity can be used in the construction of $\Omega$, affecting the contribution of each option to the loss function. In this manner, options used as benchmarks and those that are heavily traded can contribute more weight to the loss function and those options deemed less important can have their influence reduced.

### 3.2.2 Approximation of the Pricing Error

All smooth functions can be locally approximated by a polynomial (e.g., a linear or quadratic function, depending on the function and region of local approximation). Of critical importance to the method of calibration proposed here is the assumption that the pricing error function is sufficiently well-behaved for this approximation to be valid. From this local approximation, estimates of both the function and its gradient are easily obtained. These estimates are then used in an iterative procedure to find the value of the parameters that minimises the estimated sum of squared pricing errors (a Robbins-Monro-type procedure that is a stochastic analog to the method of steepest descent).

In order to fit the local polynomial, the pricing error function needs to be sampled at a number of different parameter settings. The parameter settings are chosen with the objective of producing a reasonable estimate of the function and its gradient with minimal computational effort. Since there are more coefficients to estimate, fitting a quadratic function requires evaluating the pricing error at more parameter settings than that needed to fit a linear function. Although the quadratic may provide a better fit to the function, this increased accuracy comes

at the expense of increased computational time. The analyst must decide whether the additional accuracy of the quadratic approximation justifies the increase in computational expense. That is, the increased accuracy of the quadratic function may reduce the number of iterations needed, but each iteration takes longer, as more function evaluations are required. This type of study is left for future work.

### 3.2.3 Simulation and Experimental Design

The study of experimental design is an important and useful area of statistics. Many research areas have been significantly influenced by the concepts of experimental design. From agriculture and manufacturing to chemistry and medicine, experimental design assists investigators in the design, data collection and data analysis phases of their study.

Suppose that an investigator wants to examine the effect of some controllable factors on a response variate. For example, in a chemical experiment, the factors could be temperature, pressure and amount of catalyst, while the response could be the amount of the desired chemical produced. A common objective is to determine the combination of factor levels maximising the amount of chemical produced. Experimental design concepts can be used by the investigator to choose the levels of the factors that will provide the most information in a fixed number of runs of the experiment. Clearly, the amount of information increases with the number of experimental runs, but so too does the cost of the experiment.

A computer simulation can be thought of as an experiment. When calibrating a simulated pricing model to the market prices of options, the parameters that

control the dynamics of the underlying stochastic variables are the factors in the experiment, the pricing error is the response variate and the simulation time can be thought of as the cost.

It is proposed to use a ***factorial*** or a ***fractional factorial design*** to determine the parameter settings for runs of the experiment. The reader is referred to any introductory textbook on experimental design such as Box, Hunter and Hunter (1978) or Montgomery (1997) for information on factorial and fractional factorial designs and their properties. In many circumstances these designs maximise the information obtained in a given number of experimental runs.

Typically, information is measured using some function of the covariance matrix of the estimated parameters, or the variance of a predicted response at a given factor setting. Optimal designs select factor settings that optimise these functions, such as minimising the variance of prediction. Factorial and fractional factorial designs are known to be optimal for many different response models and information functions (Box and Draper (1987)). Although these designs may be optimal at a given iteration of this calibration algorithm, we make no claim that they give a sequentially optimal design across iterations.

As an example, consider the case where three parameters control the simulation and a linear function approximates the pricing error. The experiment is to be run at the current estimate of the minimum, which will also be the centre of the experimental region, as well as at other parameter settings (i.e., at a level above and below the current estimate of the minimum). The parameters are coded so that the centre is assigned the value of zero and the higher and lower levels are

| Experimental Run | $\theta_1$ | $\theta_2$ | $\theta_3$ |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 2 | 1 | -1 | -1 |
| 3 | -1 | 1 | -1 |
| 4 | -1 | -1 | 1 |
| 5 | 0 | 0 | 0 |

Table 3.1: Layout for a fractional factorial experiment with 3 factors at 2 levels, plus a centre point.

coded plus and minus one respectively. A *layout* for this experiment is given by Table 3.1. Each column of the table corresponds to a parameter (factor) and each row corresponds to a run of the experiment. That is, the simulation is performed with that combination of parameter settings.

In this layout the third column is the product of the first two columns, resulting in a *2-level fractional factorial* design with one centre point. This layout leads to the following design matrix (in coded variables), with the column of ones corresponding to the intercept term in the linear model.

$$\tilde{x} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.3}$$

There are many other layouts that would be appropriate for this experiment. For example, a *2-level full factorial* design could have been used, but this would require

8 runs, plus the run at the centre point, effectively doubling the time required to perform the experiment. Alternatively, factorial designs with more levels for the factors could be used but this would increase the number of experimental runs. The above design is simple, requires a small number of runs and still allows for estimation of the coefficients in the linear model as described below.

### 3.2.4 Fitting the Local Polynomial

Let the index $i = 0, 1, \ldots$, indicate the number of iterations the algorithm has performed. At iteration $i$, there is a subset of the parameter space, $\Theta^i \subset \Theta$, that is the current experimental region. That is, $\Theta^i$ contains all parameter settings at which the experiment is performed on this iteration. Then, for each parameter setting, $\theta_k^i \in \Theta^i, k = 0, \ldots, K$, we perform $N$ independent replications of the experiment. In general, the number of parameter settings and the number of simulations could depend on $i$. Here, we make the simplifying assumption that these are the same for all iterations. For each replication $J$ pricing errors are produced corresponding to the set of $J$ options. This gives a multivariate response for each simulation. Common random numbers are used at each parameter setting to cut down on simulation noise. This results in a more accurate estimate of the gradient of the function. Then, using a local linear approximation to the pricing error, the

following response model is fit.

$$\begin{aligned} Y^i_{njk} &= D_{nj}(\theta^i_k) + \epsilon^i_{njk} \\ &= P_{nj}(\theta^i_k) - P_j + \epsilon^i_{njk} \\ &= \beta^i_{0,j} + \sum_{m=1}^{P} \beta^i_{m,j}\theta^i_{m,k} + \eta^i_{jk} + \epsilon^i_{njk} \\ &= \beta^i_{0,j} + \sum_{m=1}^{P} \beta^i_{m,j}\theta^i_{m,k} + \varepsilon^i_{njk}, \end{aligned} \tag{3.4}$$

where $\varepsilon^i_{njk} = \eta^i_{jk} + \epsilon^i_{njk}$, for $n = 1, \ldots, N$, $j = 1, \ldots, J$ and $k = 0, \ldots, K$. The response variate $Y^i_{njk}$ is a random variable whose distribution represents the range of possible pricing errors for option $j$ at parameter setting $k$ on the $i$th iteration of the algorithm. Here, the error terms $\eta^i_{jk}$ and $\epsilon^i_{njk}$ correspond to the (nonrandom) error due to the local linear approximation and to the (random) error from the simulation respectively. Table 3.2 provides a reference for the many symbols used in the model. To ease notation, the above equation can be written in vector form.

| | |
|---|---|
| $i =$ | iteration index |
| $p =$ | dimension of $\theta$ |
| $J =$ | total number of options |
| $j =$ | option index |
| $K =$ | number of runs at each iteration |
| $k =$ | run index |
| $N =$ | number of simulations at each iteration |
| $n =$ | simulation index |
| $\eta =$ | linear approximation error |
| $\epsilon =$ | simulation error |
| $\varepsilon =$ | total error |
| $\theta =$ | parameter settings |
| $\beta =$ | coefficients of linear model |

Table 3.2: Symbols for the linear model.

For each replication, define the $((K+1)J \times 1)$ response vector as

$$Y_n^i = \begin{pmatrix} Y_{n,1,0}^i \\ \vdots \\ Y_{n,1,K}^i \\ \vdots \\ Y_{n,J,0}^i \\ \vdots \\ Y_{n,J,K}^i \end{pmatrix}, \tag{3.5}$$

for $n = 1, \ldots, N$. Also, define the $((K+1) \times (p+1))$ matrix

$$\tilde{X}^i = \begin{pmatrix} 1 & \theta^i_{1,0} & \cdots & \theta^i_{p,0} \\ \vdots & \vdots & & \vdots \\ 1 & \theta^i_{1,K} & \cdots & \theta^i_{p,K} \end{pmatrix} \tag{3.6}$$

which serves as the building block for the $((K+1)J \times (p+1)J)$ block diagonal design matrix, $\tilde{X}^i$, defined by

$$\tilde{X}^i = \begin{pmatrix} \tilde{X}^i & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{X}^i & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \tilde{X}^i \end{pmatrix}, \tag{3.7}$$

where $\mathbf{0}$ is the $((K+1)\times(p+1))$ matrix of zeroes. Furthermore, define the coefficient vectors

$$\beta^i_j = \begin{pmatrix} \beta^i_{0,j} \\ \beta^i_{1,j} \\ \vdots \\ \beta^i_{p,j} \end{pmatrix} \quad \text{and} \quad \beta^i = \begin{pmatrix} \beta^i_1 \\ \vdots \\ \beta^i_J \end{pmatrix}, \tag{3.8}$$

for $j = 1, \ldots, J$, so that $\beta^i$ is a $((p+1)J \times 1)$ column vector. Also define the (nonrandom) linear approximation error vectors

$$\eta_j^i = \begin{pmatrix} \eta_{j,0}^i \\ \vdots \\ \eta_{j,K}^i \end{pmatrix} \quad \text{and} \quad \eta^i = \begin{pmatrix} \eta_1^i \\ \vdots \\ \eta_J^i \end{pmatrix}, \tag{3.9}$$

for $j = 1, \ldots, J$ so that $\eta^i$ is a $((K+1)J \times 1)$ column vector. Similarly, we define error vectors corresponding to the (random) simulation error. They are

$$\epsilon_{nj}^i = \begin{pmatrix} \epsilon_{nj,0}^i \\ \vdots \\ \epsilon_{nj,K}^i \end{pmatrix} \quad \text{and} \quad \epsilon_n^i = \begin{pmatrix} \epsilon_{n,1}^i \\ \vdots \\ \epsilon_{n,J}^i \end{pmatrix}, \tag{3.10}$$

for $n = 1, \ldots, N$ and $j = 1, \ldots, J$ so that $\epsilon_n^i$ is a $((K+1)J \times 1)$ column vector. Combining the error vectors $\eta^i$ and $\epsilon_n^i$, we define a $((K+1)J \times 1)$ column vector corresponding to the total error, namely

$$\varepsilon_n^i = \eta^i + \epsilon_n^i, \tag{3.11}$$

for $n = 1, \ldots, N$. Thus, for each replication the response model is

$$Y_n^i = X^i \beta^i + \varepsilon_n^i, \tag{3.12}$$

for $n = 1, \ldots, N$, where the $\varepsilon_n^i$'s are *i.i.d.* random vectors with mean $\eta^i$ and $((K+1)J \times (K+1)J)$ covariance matrix $\tilde{\Sigma}^i$. That is, the random simulation error vector $\epsilon_n^i$

has mean zero and covariance matrix $\tilde{\Sigma}^i$. The covariance of the pricing errors arises from the use of common random numbers. Now define the $((K+1)JN \times 1)$ vectors $Y^i = [(Y_1^i)', \ldots, (Y_N^i)']'$, $\varepsilon^i = [(\varepsilon_1^i)', \ldots, (\varepsilon_N^i)']'$ and the $((K+1)JN \times (p+1)J)$ matrix

$$X^i = \begin{pmatrix} \dot{X}^i \\ \vdots \\ \dot{X}^i \end{pmatrix}, \tag{3.13}$$

where $\dot{X}^i$ is repeated $N$ times. Then, the response model can be written in the familiar form,

$$Y^i = X^i \beta^i + \varepsilon^i, \tag{3.14}$$

where $E[\varepsilon^i] = \eta_N^i$ is the $((K+1)JN \times 1)$ vector constructed by stacking the vector $\eta^i$ together $N$ times and $Cov(\varepsilon^i) = \Sigma^i$, is block diagonal with $\tilde{\Sigma}^i$ repeated $N$ times along the diagonal ($\Sigma^i$ is of dimension $((K+1)JN \times (K+1)JN)$).

The estimation of the above model is easily accomplished using Weighted Least Squares (WLS). In particular, the parameter $\beta^i$ is estimated by,

$$\begin{aligned} \hat{\beta}^i &= \left(X^{i\prime}\Sigma^{i-1}X^i\right)^{-1} X^{i\prime}\Sigma^{i-1}Y^i \\ &= \left(\sum_{n=1}^{N} \dot{X}^{i\prime}\tilde{\Sigma}^{i-1}\dot{X}^i\right)^{-1} \left(\sum_{n=1}^{N} \dot{X}^{i\prime}\tilde{\Sigma}^{i-1}Y_n^i\right). \end{aligned} \tag{3.15}$$

The expectation of this estimator is

$$\begin{aligned} E\left[\hat{\beta}^{\iota}\right] &= \left(X^{\iota\prime}\Sigma^{\iota-1}X^{\iota}\right)^{-1}X^{\iota\prime}\Sigma^{\iota-1}E\left[X^{\iota}\beta^{\iota}-\varepsilon^{\iota}\right] \\ &= \beta^{\iota}+\left(X^{\iota\prime}\Sigma^{\iota-1}X^{\iota}\right)^{-1}X^{\iota\prime}\Sigma^{\iota-1}\eta_N^{\iota} \\ &= \beta^{\iota}+\left(\sum_{n=1}^{N}\dot{X}^{\iota\prime}\dot{\Sigma}^{\iota-1}\dot{X}^{\iota}\right)^{-1}\left(\sum_{n=1}^{N}\dot{X}^{\iota\prime}\dot{\Sigma}^{\iota-1}\eta^{\iota}\right). \end{aligned} \tag{3.16}$$

revealing that, because of the linear approximation error, $\hat{\beta}^{\iota}$ is a biased estimator of $\beta^{\iota}$. However, as the experimental region shrinks, the error due to the linear approximation gets small provided that $D$ is a sufficiently well-behaved function of $\theta$ and the bias in the above estimator goes to zero. That is, as $\Theta^{\iota}$ shrinks $\|\eta^{\iota}\| \to 0$ and $E[\hat{\beta}^{\iota}] \to \beta^{\iota}$, where $\|\cdot\|$ denotes the Euclidean norm. Therefore, provided that the experimental region shrinks at a certain rate from one iteration to the next, we will eventually obtain an exact local linear approximation to the function $D$ and the parameter estimates $\hat{\beta}^{\iota}$ will be unbiased.

The covariance matrix of the estimator is

$$Cov\left[\hat{\beta}^{\iota}\right] = \left(X^{\iota\prime}\Sigma^{\iota-1}X^{\iota}\right)^{-1} = \left(\sum_{n=1}^{N}\dot{X}^{\iota\prime}\dot{\Sigma}^{\iota-1}\dot{X}^{\iota}\right)^{-1}, \tag{3.17}$$

which assumes that the covariance matrix $\dot{\Sigma}^{\iota}$ is known. For a large number of simulations the sample covariance matrix $\widehat{\dot{\Sigma}^{\iota}_N} \approx \dot{\Sigma}^{\iota}$. In our case $N$ is large and the covariance matrix is well-estimated by

$$\widehat{Cov}\left[\hat{\beta}^{\iota}\right] = \left(\sum_{n=1}^{N}\dot{X}^{\iota\prime}\left(\widehat{\dot{\Sigma}^{\iota}_N}\right)^{-1}\dot{X}^{\iota}\right)^{-1}. \tag{3.18}$$

For a parameter setting $\tilde{\theta} \in \Theta^i$, define the $(J \times (p+1)J)$ design matrix $A(\tilde{\theta})$ as

$$A(\tilde{\theta}) = \begin{pmatrix} (1\ \tilde{\theta}') & \tilde{\mathbf{0}} & \tilde{\mathbf{0}} & \cdots & \tilde{\mathbf{0}} \\ \tilde{\mathbf{0}} & (1\ \tilde{\theta}') & \tilde{\mathbf{0}} & \cdots & \tilde{\mathbf{0}} \\ \vdots & & \ddots & & \vdots \\ \tilde{\mathbf{0}} & \tilde{\mathbf{0}} & \cdots & \tilde{\mathbf{0}} & (1\ \tilde{\theta}') \end{pmatrix}, \tag{3.19}$$

where $(1\ \tilde{\theta}')$ is repeated $J$ times and $\tilde{\mathbf{0}}$ is the $(1 \times (p+1))$ row vector of zeros. Then, the predicted mean pricing error at $\tilde{\theta}$ is

$$\hat{D}^i(\tilde{\theta}) = A(\tilde{\theta})\hat{\beta}^i. \tag{3.20}$$

The $(J \times J)$ covariance matrix corresponding to this estimator is

$$Cov(\hat{D}^i(\tilde{\theta})) = A(\tilde{\theta}) \left( X^{i\prime} \Sigma^{i-1} X^i \right)^{-1} A(\tilde{\theta})' \tag{3.21}$$

which can be estimated in the obvious way.

### 3.2.5 Updating the Estimated Optimum

At iteration $i$ we have a current estimate of the minimum, $\hat{\theta}_0^i \in \Theta^i$. Also suppose that the simulation has been performed and the coefficients of the approximating polynomial have been estimated as in Section 3.2.4. Again, for illustration purposes the approximating polynomial is a linear function of $\theta$. Extension of the updating scheme to the quadratic case is straightforward.

The task is to find the value of $\hat{\theta}_0^{i+1} \in \Theta^i$ that minimises the SSPE. The fitted

pricing error obtained from estimating the linear model $\hat{D}^i(\theta)$ is used as an approximation to $D(\theta)$. A search direction for a new estimate of the minimum is given by the gradient of the fitted SSPE at $\hat{\theta}_0^i$. Specifically the gradient is

$$g^i(\hat{\theta}_0^i) = 2\hat{D}^i(\hat{\theta}_0^i)'\nabla\hat{D}^i(\hat{\theta}_0^i), \tag{3.22}$$

where the $j^{th}$ row of $\nabla\hat{D}^i$ is $\left(\hat{\beta}_{1,J}^i \;\; \hat{\beta}_{2,J}^i \;\; \cdots \;\; \hat{\beta}_{p,J}^i\right)$ for $j = 1, \ldots, J$.

Since $g^i$ is computed using the estimated linear approximation for $D$, there is no guarantee that values of $\theta$ in the direction $g^i$ from $\hat{\theta}_0^i$ will have an actual SSPE that is less than SSPE$(\hat{\theta}_0^i)$. However, the fitted SSPE will be less for these values of $\theta$. For this reason we accept a new candidate for the minimum $(\hat{\theta}_{0,c}^i)$ as the new estimate of the optimum only if its SSPE is less than SSPE$(\hat{\theta}_0^i)$, where the SSPE's are estimated with the same random number stream. Candidates for the new estimated minimum are computed using $g^i$ as the direction in a line search method combining step-halving and alternating directions. The details of this procedure, as well as the updating of the experimental region, are described in the program below.

1. Compute a candidate for the new minimum by

$$\hat{\theta}_{0,c}^i = \hat{\theta}_0^i - \delta^i \frac{g^i(\hat{\theta}_0^i)}{\left\| g^i(\hat{\theta}_0^i) \right\|}, \tag{3.23}$$

where $\delta^i$ is the current step size. The step size can be either a scalar or vector and is chosen to ensure $\hat{\theta}_{0,c}^i \in \Theta^i$.

2. With the same random number stream, run the simulation at $\hat{\theta}^i_{0,c}$ and estimate $\text{SSPE}(\hat{\theta}^i_{0,c})$.

3. If $\text{SSPE}(\hat{\theta}^i_{0,c}) \leq \text{SSPE}(\hat{\theta}^i_0)$

   (a) set $\hat{\theta}^{i+1}_0 = \hat{\theta}^i_{0,c}$ and go to Step 5(a).

4. Else let $\tilde{\delta}^i = \delta^i$ and $h = 0$. Then

   (a) let $\tilde{\delta}^i = -\tilde{\delta}^i$ (to reverse direction) and calculate a new candidate for the minimum by

$$\hat{\theta}^i_{0,c1} = \hat{\theta}^i_0 - \tilde{\delta}^i \frac{g^i(\hat{\theta}^i_0)}{\left\| g^i(\hat{\theta}^i_0) \right\|}. \tag{3.24}$$

   (b) With the same random number stream, run the simulation at $\hat{\theta}^i_{0,c1}$ and estimate $\text{SSPE}(\hat{\theta}^i_{0,c1})$.

   (c) If $\text{SSPE}(\hat{\theta}^i_{0,c1}) \leq \text{SSPE}(\hat{\theta}^i_0)$

      i. set $\hat{\theta}^{i+1}_0 = \hat{\theta}^i_{0,c1}$ and go to Step 5.

   (d) Else

      i. let $h = h + 1$, $\tilde{\delta}^i = -0.5\tilde{\delta}^i$ (to step-half and reverse direction) and calculate a new candidate for the minimum by

$$\hat{\theta}^i_{0,c2} = \hat{\theta}^i_0 - \tilde{\delta}^i \frac{g^i(\hat{\theta}^i_0)}{\left\| g^i(\hat{\theta}^i_0) \right\|}. \tag{3.25}$$

      ii. With the same random number stream, run the simulation at $\hat{\theta}^i_{0,c2}$ and estimate $\text{SSPE}(\hat{\theta}^i_{0,c2})$.

iii. If $\text{SSPE}(\hat{\theta}^i_{0.c2}) \leq \text{SSPE}(\hat{\theta}^i_0)$

A. set $\hat{\theta}^{i+1}_0 = \hat{\theta}^i_{0.c2}$ and go to Step 5.

iv. Else if $h = H$ ($H$ is the maximum number of times step-halving is allowed)

A. declare $\hat{\theta}^i_0$ a local minimum and terminate the algorithm.

v. Else

A. go to Step 4(a) with the current step size $\tilde{\delta}^i$.

5. If $h = 0$ (no step-halving was performed in Step 4),

(a) set $\delta^{i+1} = \delta^i$ and let $\Theta^{i+1}$ be the translation of $\Theta^i$ having $\hat{\theta}^{i+1}_0$ at its centre, keeping the size of the experimental region the same.

6. Else ($h > 0$ and step-halving was performed in Step 4)

(a) curvature is suspected or random error larger than the decrease in direction of the gradient is present and we shrink the experimental region to help ensure that the local linear approximation is valid. Thus, first translate $\Theta^i$ so that $\hat{\theta}^{i+1}_0$ is at its centre, then shrink the region by muliplying its boundary coordinates by a positive constant $a < 1$. This new region defines $\Theta^{i+1}$ and set $\delta^{i+1} = a\delta^i$.

Note that one can transform the parameters so that they all lie on the same scale, using the logistic transform, for example. This transform is employed in the examples that follow and makes the search for the optimal parameter settings easier, since the transformed parameter space is $\Re^p$ as opposed to some bounded subset of $\Re^p$.

### 3.2.6 Calibration Algorithm

Below are the steps of the proposed calibration algorithm, followed by a short discussion.

1. Set $i = 0$ and choose a starting value for the estimated minimum $\hat{\theta}_0^i$. Then, define an initial experimental region $\Theta^i \subset \Theta$ with centre $\theta_0^i = \hat{\theta}_0^i$ and also define an appropriate initial step size $\delta^i$.

2. Choose a set of $K$ parameter settings such that $\theta_k^i \in \Theta^i, k = 1, \ldots, K$, for a total of $K + 1$ parameter settings including $\theta_0^i$ (see Section 3.2.3).

3. Using common random numbers, estimate (by simulation) the pricing error, $D_j(\theta_k^i)$ for option $j$ at parameter setting $k$, $j = 1, \ldots, J$, $k = 0, \ldots, K$.

4. Fit the approximating polynomial as in Section 3.2.4.

5. Update the estimated minimum and experimental region as in Section 3.2.5.

6. Let $i = i + 1$ and iterate Steps 2-6.

This algorithm is a Robbins-Monro-type stochastic approximation procedure. Robbins-Monro procedures and their variants have been extensively studied and the interested reader is referred to Benveniste et al (1990) and Ermoliev and Wets (1988) for more information on the properties of estimators derived from these procedures.

It is important to note that this algorithm does not ensure that a global minimum is achieved. In fact, for a finite number of iterations, and due to the simulation noise it is unclear if an exact local minimum is achieved. However, the estimated

minimum should be close to a local minimum. Keeping in mind the objective of this algorithm, neither problem poses a serious concern. The goal is to determine a parameter setting that yields option prices within the bid/ask spread. In fact, there may be many parameter settings able to achieve this objective. If one such setting is produced, then the pricing model is calibrated at this setting and the algorithm is deemed successful (along with the pricing model). The theory of option pricing in incomplete markets asserts that there may be many risk-neutral measures consistent with a set of options. Thus, the non-uniqueness of the calibrating parameters does not pose a problem from a theoretical finance perspective. Each parameter setting that calibrates the model corresponds to a different risk-neutral measure. That is, there exists $\Theta^* \subset \Theta$ such that for all $\theta \in \Theta^*$ the model prices computed using $\theta$ are consistent with the observed market prices.

When calibrating a certain pricing model to a set of option prices there may be no parameter settings that price all options within the bid/ask spread. However, as long as the SSPE is minimised, the model is considered calibrated. We take the average of the bid and ask prices as the market price an option. The fact that the calibrated model incorrectly prices some options is due to a failure in the underlying model, rather than the calibration algorithm. An obvious example of a model's inability to correctly price all options, yet still be considered calibrated is given by the Black-Scholes model. Typically there is no single value of the volatility of the underlying asset that yields correct prices for all options at a single maturity. The volatility smile/smirk is a well-documented phenomenon. However, for a given maturity, analysts can compute a Black-Scholes calibrating volatility that minimises

the SSPE (see Equation 3.34).

## 3.3 Calibration Examples: Model Option Prices

### 3.3.1 Black-Scholes Model

The Black-Scholes model assumes that the price of the underlying asset evolves according to geometric Brownian motion (GBM) with constant drift $\mu$ and constant diffusion $\sigma$ (Black and Scholes (1973)). After a change of measure, the risk-neutral evolution of the asset is described by

$$dS_t = rS_t dt + \sigma S_t dW_t, \tag{3.26}$$

where $r$ is the risk-free rate of interest and $W_t$ is a risk-neutral standard Brownian motion.

This setup leads to the famous Black-Scholes pricing formula for European call options, namely

$$BS(S_0, r, T, X, \sigma) = S_0 N(d_1) - Xe^{-rT} N(d_2), \tag{3.27}$$

where $S_0$ is the current price, $T$ is the time until maturity, $X$ is the strike price, $d_1 = \frac{\ln(S_0/X)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$, $d_2 = d_1 - \sigma\sqrt{T}$, and $N(\cdot)$ is the standard normal cumulative distribution function (CDF).

From the market one can obtain $S_0$, $r$ and the price of the option. Obviously, the strike price and time to maturity are also available. The only information

| Strike | Option Price | Strike | Option Price |
|---|---|---|---|
| 80 | 22.05 | 110 | 2.13 |
| 85 | 17.37 | 115 | 1.14 |
| 90 | 13.01 | 120 | 0.57 |
| 95 | 9.17 | 125 | 0.27 |
| 100 | 6.05 | 130 | 0.12 |
| 105 | 3.72 | 135 | 0.05 |

Table 3.3: Black-Scholes model call option prices.

not directly observable is the diffusion parameter $\sigma$. Typically practitioners invert the Black-Scholes formula and quote an option price in terms of its Black-Scholes implied volatility. This is the volatility that makes the model price match the market price. In other words, $\sigma$ is used as a calibrating parameter.

Since an analytic form for the option price exists (Equation 3.27), simulation is obviously not required to price options using this model. However, simulation can be used to value options with this model and, since exact solutions are available, this framework provides a simple illustration of the proposed calibration algorithm.

In this example, call option prices at a single maturity are generated from Equation 3.27 with $S_0 = 100$, $r = .05$ and $T = 0.5$ years. The volatility is $\sigma = 0.1693$ and was chosen at random from the interval [0.1,0.4] using the continuous uniform distribution defined on this interval. The option prices, along with the strike prices are given in Table 3.3. These option prices are used as inputs to the calibration procedure. The starting value for the calibration algorithm is $\hat{\sigma}_0 = 0.25$ (the centre of the range of possible values). Simulations are run at 3 levels of volatility which can be coded (+1,-1,0). At each iteration and for each level of volatility 20,000 simulations are performed.

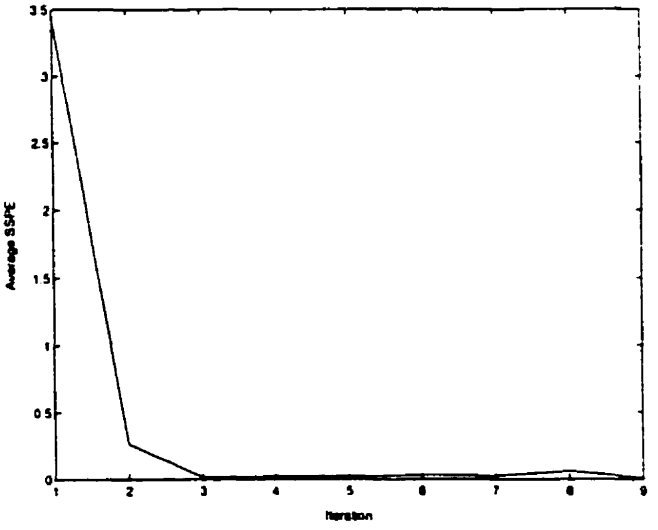


Figure 3.1: Average SSPE versus iteration number for the Black-Scholes model calibration to model-produced options.

Two simulations are performed at $\hat{\sigma}^i$, one at iteration $i-1$ and another at iteration $i$. Corresponding to the different random number streams used at iteration $i-1$ and $i$, two unbiased estimates of the SSPE at this parameter setting are available. Figure 3.1 plots the average SSPE (of these two estimates) against the iteration number to illustrate the convergence of the algorithm.

Using the estimate after the last iteration as the calibrating value ($\hat{\sigma} = 0.1690$), option prices are computed and compared with the true option prices in Table 3.4. It is obvious from these results that the calibration algorithm is quite successful here. As an alternative to using the estimated parameter value at the last (10th) iteration, one could average $\hat{\sigma}^i$ across iterations that give low values of SSPE. From Figure 3.1 it is observed that the average SSPE is small for iterations 3 to 10. Averaging over these values gives $\hat{\sigma} = 0.16934$, which is equal to the true value to

| Strike | Calibrated Option Price | Actual Option Price |
|---|---|---|
| 80 | 22.05 | 22.05 |
| 85 | 17.37 | 17.37 |
| 90 | 13.00 | 13.01 |
| 95 | 9.17 | 9.17 |
| 100 | 6.04 | 6.05 |
| 105 | 3.71 | 3.72 |
| 110 | 2.13 | 2.13 |
| 115 | 1.14 | 1.14 |
| 120 | 0.57 | 0.57 |
| 125 | 0.27 | 0.27 |
| 130 | 0.12 | 0.12 |
| 135 | 0.05 | 0.05 |

Table 3.4: Calibrated Black-Scholes option prices computed using $\hat{\sigma} = 0.1690$ along with the actual prices computed using the true volatility $\sigma = 0.1693$.

4 decimal places.

### 3.3.2 Hull-White Model

The Hull-White model is a stochastic volatility model where the variance of the stock price follows its own stochastic process. Specifically, in a risk-neutral world a stock price ($S$) and its instantaneous variance ($\sigma^2$) follow the stochastic processes

$$dS_t = rS_t dt + \sigma_t S_t dW_t \tag{3.28}$$

and

$$d\sigma_t^2 = \mu dt + \xi \sigma_t^2 dZ_t, \tag{3.29}$$

where $r$ is the risk-free rate and $W$ and $Z$ are Wiener processes with instantaneous correlation $\rho$. The parameters $\mu$ and $\xi$ may depend on $S$ as well as $\sigma$ and $t$. Assume that $\sigma^2$ is uncorrelated with aggregate consumption so that risk-neutral valuation can be used (see Hull and White (1987) for more details).

In general, option prices based on this model must be computed via simulation. Using $M$ equally-spaced time intervals of length $\Delta t$ (i.e., $t = M\Delta t$), the simulation of $S$ and $\sigma$ is performed by

$$S_m = S_{m-1} \exp\left[(r - \sigma_{m-1}^2/2)\Delta t + u_m \sqrt{\sigma_{m-1}^2 \Delta t}\right] \tag{3.30}$$

and

$$\sigma_m^2 = \sigma_{m-1}^2 \exp\left[(\mu - \xi^2/2)\Delta t + \rho u_m \xi \sqrt{\Delta t} + v_m \xi \sqrt{1-\rho^2}\sqrt{\Delta t}\right], \tag{3.31}$$

where $S_m$ is the stock price at time $m\Delta t$, $\sigma_m^2$ is the variance at time $m\Delta t$ and $u_m$ and $v_m$ are independent standard normal random variables for $m = 1, \ldots, M$. Note that the values of $\mu$ and $\xi$ are updated using $\sigma_{m-1}^2$ and $S_{m-1}$.

The calibrating parameters for this model are the drift and diffusion ($\mu$ and $\xi$) of the stochastic process for the variance, the correlation between the stock price and variance processes ($\rho$) and the initial variance ($\sigma_0^2$) used in the simulation. The functional forms of $\mu$ and $\xi$ dictate the total number of calibrating parameters.

### Zero-drift Variance Process

Here the drift of the variance process is set equal to zero ($\mu = 0$) and $\xi$ is assumed to be constant so that three parameters require calibration ($\sigma_0^2, \rho$ and $\xi$). It is assumed that $\sigma_0^2 \in [0.01, 0.17]$ and $\rho \in [-1, 1]$. From Equation 3.31 and with $\mu = 0$ it is easily seen that

$$\ln\left(\frac{\sigma_m^2}{\sigma_{m-1}^2}\right) \sim N\left(-\frac{\xi^2}{2}\Delta t,\ \xi^2 \Delta t\right). \tag{3.32}$$

A plausible range of values for $\xi$ is determined using a probabilistic argument. If we constrain the absolute percentage change in volatility ($|\frac{\sigma_m - \sigma_{m-1}}{\sigma_{m-1}}|$) to be less than 10% for nearly every time step, then this translates to a probability statement such as

$$P\left(0.9 < \sqrt{\frac{\sigma_m^2}{\sigma_{m-1}^2}} < 1.1\right) = 0.95 \tag{3.33}$$

from which a reasonable value of $\xi$ can be easily computed. Solving the above equation gives $\xi = 1.6$. Since $\xi > 0$, a reasonable constraint on $\xi$ is that $\xi \in [0.1, 3.1]$.

With $S_0 = 1336.43$, $\Delta t = \frac{1}{252}$, $r = 0.04$, $\sigma_0^2 = 0.1326$, $\rho = -0.5942$ and $\xi = 0.7934$, the prices of European call options covering a range of strike prices with maturities of 20 and 40 days were computed using 50,000 simulations. The above parameter values were chosen at random using the continuous uniform distribution defined on the appropriate ranges. These prices, given in Table 3.5, are used as the known market prices in the calibration algorithm. We ignore the simulation error present in these prices.

The starting value for each calibrating parameter is the center of its range of possible values. At each iteration the simulation is run at five different parameter combinations according to the layout given in Table 3.1. Twenty thousand replications are used at each combination. The width of the current experimental region ($\Theta^i$) has an effect on the quality of the linear approximation as well as how much the estimate of the minimum moves at each iteration. Care must be taken to select a width for the algorithm to proceed as expected.

Figure 3.2 shows the average SSPE versus the iteration number illustrating the convergence of the algorithm. As before the average SSPE is computed by averaging the two estimates of SSPE($\hat{\theta}^i$) obtained with the different random number streams used at iterations $i-1$ and $i$. From the final iteration the calibrated values of the parameters are $\hat{\sigma}_0^2 = 0.1296$, $\hat{\rho} = -0.2711$ and $\hat{\xi} = 0.9031$. Figures 3.3 and 3.4 show point estimates of the pricing errors along with approximate pointwise

| Strike | 20-day Option Price | 40-day Option Price |
|---|---|---|
| 1100 | — | 251.44 |
| 1150 | — | 208.02 |
| 1200 | 150.16 | 167.90 |
| 1225 | 129.61 | — |
| 1250 | 110.40 | 131.79 |
| 1275 | 92.68 | — |
| 1300 | 76.65 | 100.38 |
| 1325 | 62.37 | — |
| 1350 | 49.91 | 74.04 |
| 1375 | 39.27 | — |
| 1400 | 30.35 | 52.82 |
| 1425 | 23.03 | — |
| 1450 | 17.12 | 36.34 |
| 1475 | 12.46 | — |
| 1500 | 8.90 | 24.12 |
| 1550 | — | 15.42 |
| 1600 | — | 9.54 |

Table 3.5: Call option prices computed from the zero-drift Hull-White Model with $\sigma_0^2 = 0.1326$, $\rho = -0.5942$, $\xi = 0.7934$ based on 50,000 simulations (— indicates no option for that strike).

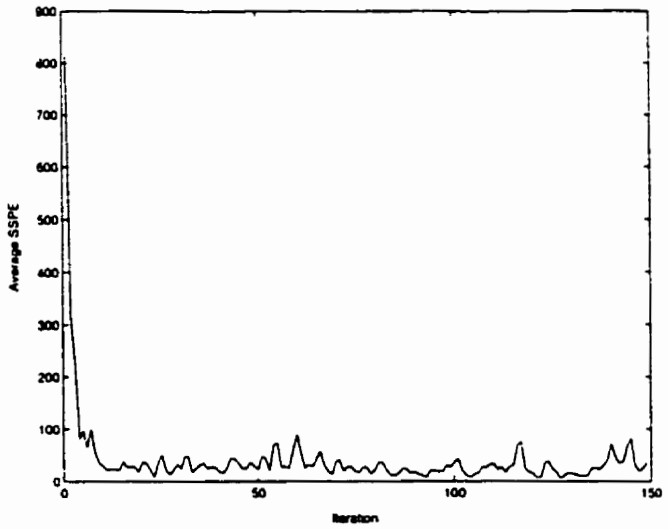


Figure 3.2: Average SSPE versus iteration number for the zero-drift Hull-White model calibration to model-produced 20-day and 40-day options.

95% confidence intervals computed using the calibrated parameters for the two maturities. Note that these results are based on the same simulation. Since zero is within the confidence interval of the pricing error at all strikes and both maturities, the calibration algorithm is successful at calibrating the zero-drift Hull-White model to simulated option prices at two maturities.

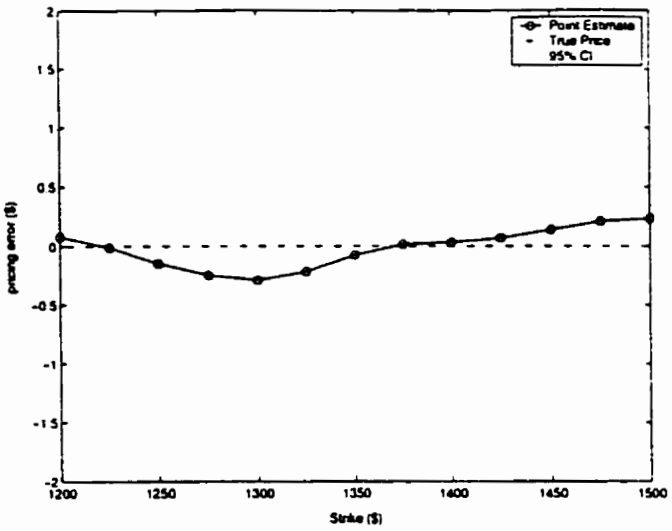


Figure 3.3: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of the 20-day zero-drift Hull-White model call options using $\hat{\sigma}_0^2 = 0.1296$, $\hat{\rho} = -0.2711$ and $\hat{\xi} = 0.9031$ based on 20,000 simulations.

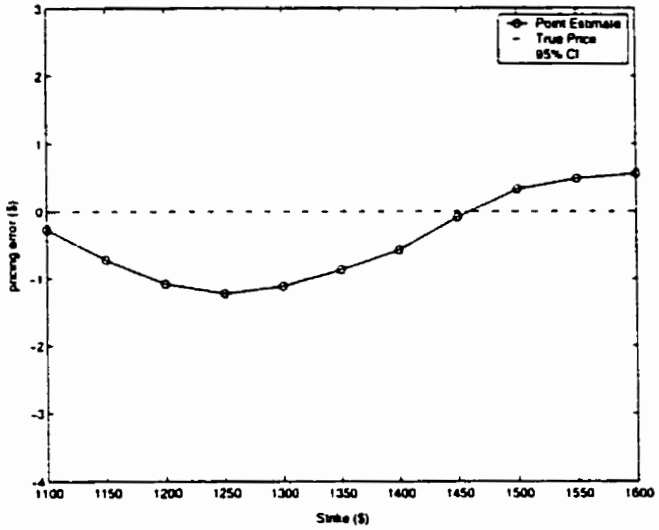


Figure 3.4: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of the 40-day zero-drift Hull-White model call options using $\hat{\sigma}_0^2 = 0.1296$, $\hat{\rho} = -0.2711$ and $\hat{\xi} = 0.9031$ based on 20,000 simulations.

### Mean-reverting Variance Process

In this example, the variance is assumed to follow a mean-reverting process and $\xi$ is constant. Specifically, $\mu = \kappa(\bar{\sigma}^2 - \sigma^2)$ where $\kappa$ is the rate of mean reversion and $\bar{\sigma}^2$ is long-run mean of the variance. This model has five parameters ($\sigma_0^2, \rho, \xi, \kappa$ and $\bar{\sigma}^2$) that could be used for calibration. However, $\bar{\sigma}^2$ can be set at a reasonable level such as the historically-observed variance or a Black-Scholes calibrating volatility, leaving four calibrating parameters.

As in the above example, $\sigma_0^2 \in [0.01, 0.17]$, $\rho \in [-1, 1]$ and $\xi \in [0.1, 3.1]$. It is also assumed that $\kappa \in [0.1, 5]$ and $\bar{\sigma}^2 = 0.04$, so that $\sigma_0^2$, $\rho$, $\xi$ and $\kappa$ are the calibrating parameters. The prices of European call options covering a range of strike prices with maturities of 20 and 40 days are computed using 50,000 simulations. The parameter values $\sigma_0^2 = 0.1414$, $\rho = -0.1106$, $\xi = 1.3671$ and $\kappa = 3.081$ were randomly chosen from the appropriate continuous uniform distributions. These parameter values along with $S_0 = 1336.43$, $\Delta t = \frac{1}{252}$, and $r = 0.04$ generated the option prices given in Table 3.6. These are used as the known market prices in the calibration algorithm. The simulation error in these prices is ignored.

The starting value for each calibrating parameter is the center of its range of possible values. At each iteration the simulation is run at nine different parameter combinations according to the layout given in Table 3.7. Twenty thousand replications are used with each combination.

The average SSPE plotted against the iteration number in Figure 3.5 shows the SSPE converges to a small value. From the final iteration, the calibrated parameter values are $\hat{\sigma}_0^2 = 0.1400$, $\hat{\rho} = -0.2073$, $\hat{\xi} = 0.7425$ and $\hat{\kappa} = 2.0603$. Figures 3.6

| Strike | 20-day Option Price | 40-day Option Price |
|---|---|---|
| 1100 | — | 252.24 |
| 1150 | — | 208.80 |
| 1200 | 151.30 | 168.78 |
| 1225 | 130.85 | — |
| 1250 | 111.75 | 132.95 |
| 1275 | 94.13 | — |
| 1300 | 78.19 | 101.92 |
| 1325 | 64.01 | — |
| 1350 | 51.65 | 76.07 |
| 1375 | 41.10 | — |
| 1400 | 32.23 | 55.23 |
| 1425 | 24.90 | — |
| 1450 | 18.97 | 39.10 |
| 1475 | 14.27 | — |
| 1500 | 10.57 | 27.06 |
| 1550 | — | 18.32 |
| 1600 | — | 12.19 |

Table 3.6: Call option prices computed from the mean-reverting Hull-White Model with $\sigma_0^2 = 0.1414$, $\rho = -0.1106$, $\xi = 1.3671$, $\kappa = 3.081$ based on 50,000 simulations (— indicates no option for that strike).

| Experimental Run | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | -1 | -1 |
| 3 | 1 | -1 | 1 | -1 |
| 4 | 1 | -1 | -1 | 1 |
| 5 | -1 | 1 | 1 | -1 |
| 6 | -1 | 1 | -1 | 1 |
| 7 | -1 | -1 | 1 | 1 |
| 8 | -1 | -1 | -1 | -1 |
| 9 | 0 | 0 | 0 | 0 |

Table 3.7: Layout for a fractional factorial experiment with 4 factors at 2 levels plus a centre point.

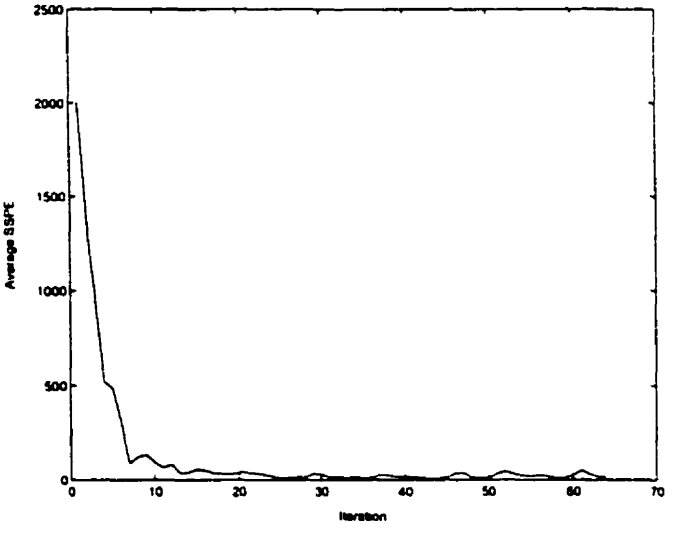


Figure 3.5: Average SSPE versus iteration number for the mean-reverting Hull-White model calibration to model-produced 20-day and 40-day options.

and 3.7 show the estimated pricing errors along with approximate pointwise 95% confidence intervals computed using the calibrated parameters for the two maturities. Note that these results are based on the same simulation. The calibration algorithm has clearly achieved its goal, as zero is within the confidence interval of the pricing error at all strikes and both maturities.

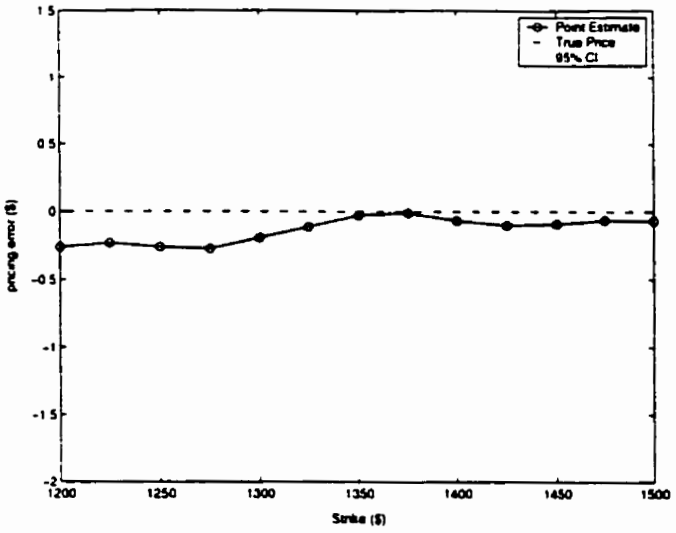


Figure 3.6: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of the 20-day mean-reverting Hull-White model call options using $\hat{\sigma}_0^2 = 0.1400$, $\hat{\rho} = -0.2073$, $\hat{\xi} = 0.7425$ and $\hat{\kappa} = 2.0603$ based on 20,000 simulations.

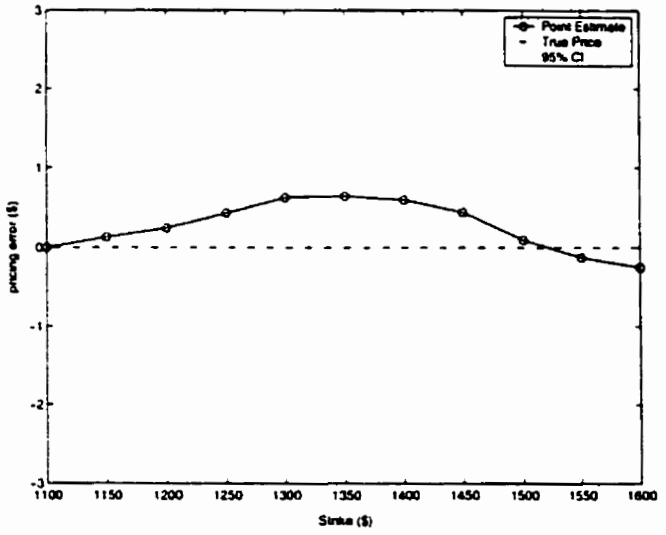


Figure 3.7: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of the 40-day mean-reverting Hull-White model call options using $\hat{\sigma}_0^2 = 0.1400$, $\hat{\rho} = -0.2073$, $\hat{\xi} = 0.7425$ and $\hat{\kappa} = 2.0603$ based on 20,000 simulations.

## 3.4 Calibration Examples: S&P500 Options

### 3.4.1 Black-Scholes Model

In this example, the Black-Scholes model is calibrated to the market prices of S&P500 call options. Calibration is performed by numerically solving for an exact value of $\sigma_{cal}$ (the usual method for Black-Scholes) as well as with the calibration algorithm presented in this paper. The results show that the simulation-based calibration method produces nearly the same value for the calibrating volatility as the exact method. However, neither method reproduces the market prices of the options revealing the failure of the Black-Scholes model.

The option prices are the average of the bid/ask spread of traded S&P500 call options at 3:30pm on May 10, 1999. The closing value of the index on this day is 1336.43. Here, only call options that expire in June, 1999 are considered, so the time to maturity of the options is $\frac{29}{252}$ years. The risk-free rate is 5% which is taken from the nearest expiring T-bill. Table 3.8 gives the prices of the June options, as well as those expiring in July which are used in Section 3.4.3.

The exact method of calibration computes the volatility ($\sigma_{cal}$) that solves

$$\sigma_{cal} = \arg\min_{\sigma>0} \sum_{j=1}^{J} [P_j(\sigma) - P_j]^2, \tag{3.34}$$

where $P_j(\sigma)$ and $P_j$ are the Black-Scholes and market prices of option $j$ respectively. For this set of options $\sigma_{cal} = 0.2213$ and this corresponds to $SSPE(\sigma_{cal}) = 240.47$.

Figure 3.8 reveals that the average SSPE converges very quickly to a neighbourhood of the true value (240.47) and then randomly bounces around in this

| Strike | June Option Price | July Option Price |
|---|---|---|
| 1100 | — | 250.88 |
| 1125 | 218.88 | 227.88 |
| 1150 | 195.25 | 205.25 |
| 1175 | 172.00 | 183.25 |
| 1190 | 158.75 | — |
| 1195 | 154.13 | — |
| 1200 | 149.50 | 161.75 |
| 1215 | 136.25 | — |
| 1225 | 127.38 | 141.00 |
| 1250 | 106.25 | 120.88 |
| 1275 | 86.25 | 101.88 |
| 1285 | 78.50 | — |
| 1290 | 74.88 | — |
| 1300 | 67.88 | 84.00 |
| 1315 | 57.50 | — |
| 1320 | 54.25 | — |
| 1325 | 51.00 | 67.63 |
| 1330 | 47.75 | — |
| 1340 | 41.88 | — |
| 1350 | 36.13 | 52.75 |
| 1355 | 33.63 | — |
| 1360 | 31.00 | — |
| 1365 | 28.50 | — |
| 1370 | 26.25 | — |
| 1375 | 24.00 | 39.50 |
| 1400 | 14.50 | 28.38 |
| 1425 | 8.13 | 19.63 |
| 1450 | 4.13 | 12.88 |
| 1475 | 1.88 | 8.00 |
| 1500 | — | 4.75 |

Table 3.8: S&P500 call option prices for June (29-day) and July (50-day) maturities (— indicates no option for that strike).

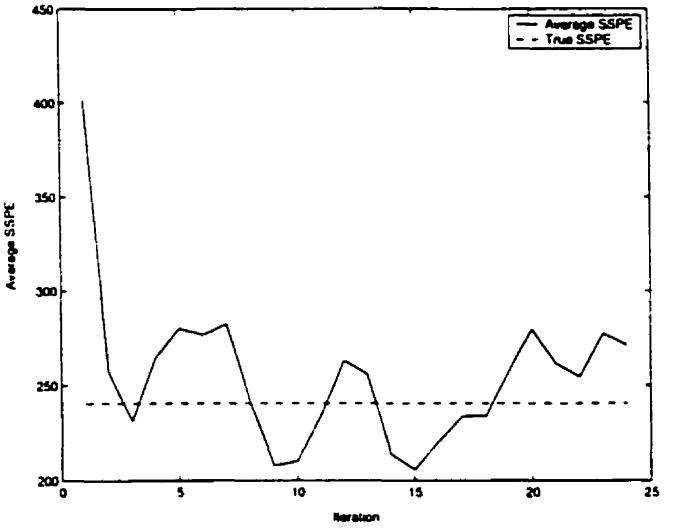


Figure 3.8: Average SSPE versus iteration number for the Black-Scholes model calibration to 29-day S&P500 options.

neighbourhood. In fact, after only 2 iterations a reasonable estimate of the calibrating volatility is produced ($\hat{\sigma} = 0.2119$). By averaging $\hat{\sigma}^i$ over iterations 2 to 25, an estimate of the calibrating volatility equal to the true value to three decimal places is $\hat{\sigma}_{cal} = 0.2210$. Figure 3.9 shows point estimates of the pricing errors as well as approximate pointwise 95% confidence intervals based on 200,000 simulations and using $\hat{\sigma} = 0.2210$.

Not surprisingly, there is significant mispricing of options that are not at-the-money. As previously mentioned, this is evidence of a failure of the Black-Scholes model, not a failure of the simulation-based calibration procedure.

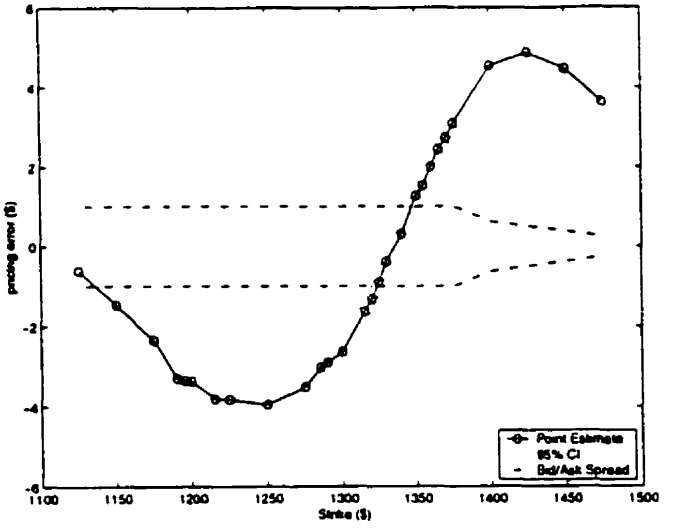


Figure 3.9: Point estimates and approximate pointwise 95% confidence intervals of the Black-Scholes pricing error of 29-day S&P500 call options using $\hat{\sigma} = 0.2210$ and based on 200,000 simulations.

### 3.4.2 REB Models

The REB models of Chapter 2 are calibrated to the 29-day options. For all three models we use both 1-day and 5-day time intervals. The options are treated as 30-day options when using 5-day time intervals, ignoring the effect of the extra day. Figures 3.10, 3.12 and 3.14 show that for all three models and with both time intervals, the average SSPE decreases as the iteration number increases, clearly showing the success of the calibration algorithm. The values of the calibrating volatilities are given in Table 2.1. Figures 3.11, 3.13 and 3.15 give point estimates and approximate pointwise 95% confidence intervals based on 200,000 simulations for REB-I, REB-L1 and REB-L2 respectively. As discussed in Chapter 2, REB models with 1-day time intervals are not appropriate for these options. However,

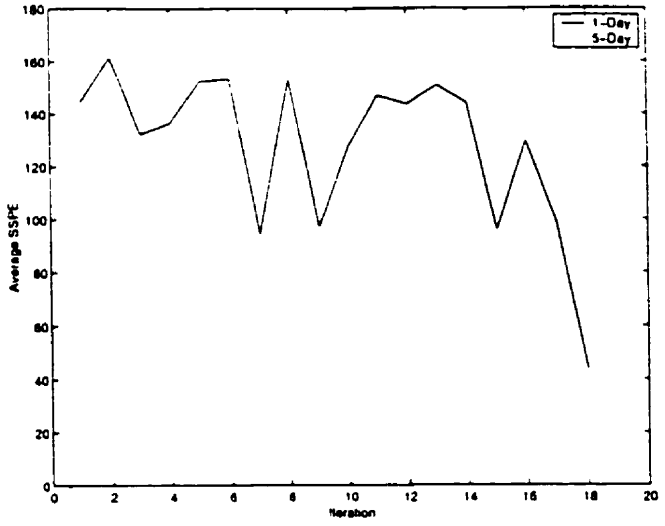


Figure 3.10: Average SSPE versus iteration number for REB-I calibration to 29-day S&P500 options.

REB models with 5-day time intervals are much more consistent with all of the options.

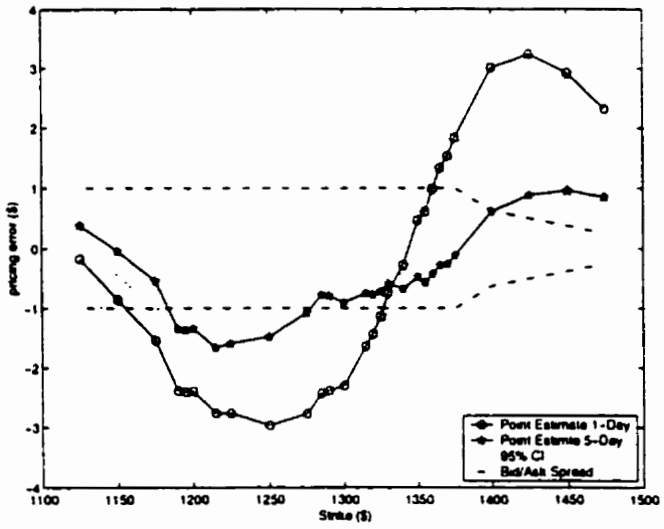


Figure 3.11: Point estimates and approximate pointwise 95% confidence intervals of the REB-I pricing error of 29-day S&P500 call options based on 200,000 simulations.

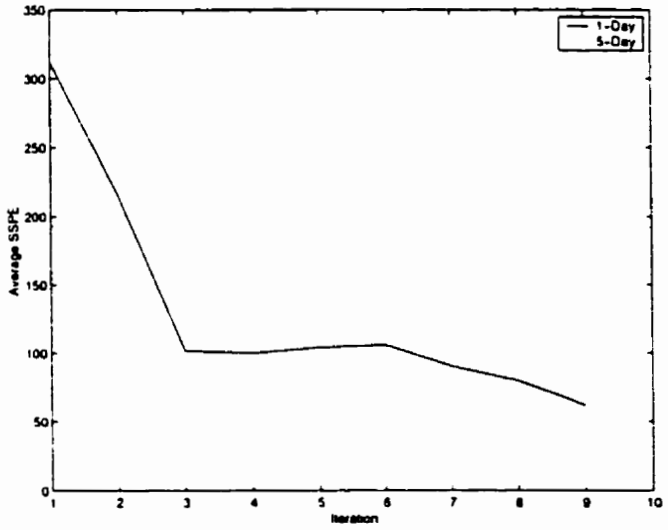


Figure 3.12: Average SSPE versus iteration number for REB-L1 calibration to 29-day S&P500 options.

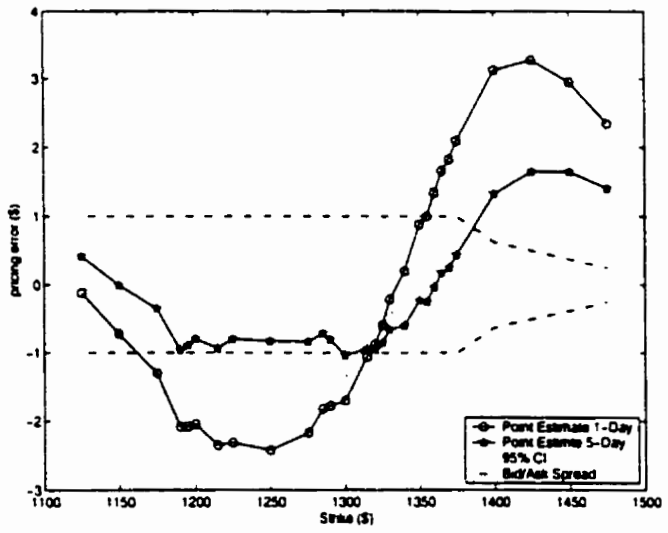


Figure 3.13: Point estimates and approximate pointwise 95% confidence intervals of the REB-L1 pricing error of 29-day S&P500 call options based on 200,000 simulations.

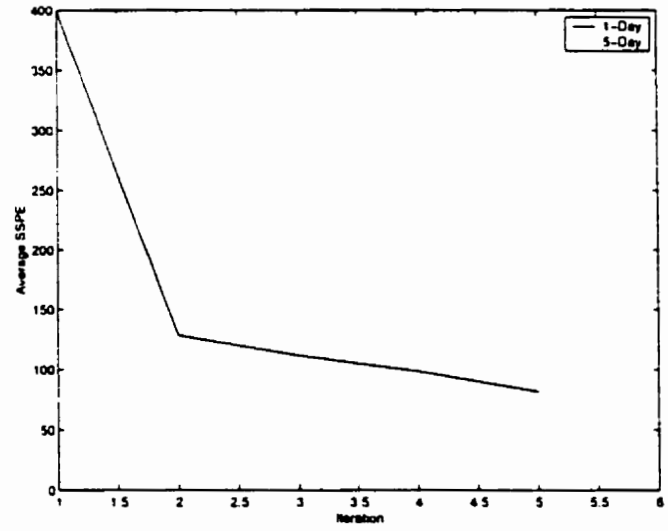


Figure 3.14: Average SSPE versus iteration number for REB-L2 calibration to 29-day S&P500 options.

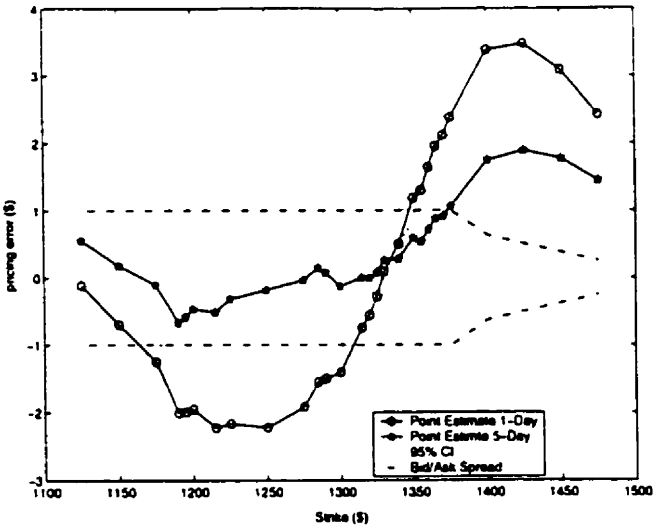


Figure 3.15: Point estimates and approximate pointwise 95% confidence intervals of the REB-L2 pricing error of 29-day S&P500 call options based on 200,000 simulations.

### 3.4.3 Hull-White Model

Here, the Hull-White model with a zero-drift variance process is used as the underlying risk-neutral data generating process. First, this model is calibrated only to the 29-day options used in the previous example. The calibration algorithm stopped after 28 iterations as a local minimum was found (i.e., step-halving occurred the maximum number of times at the final iteration). Figure 3.16 shows the average SSPE decreases as the iteration number increases. The calibrated parameter values are $\hat{\sigma}_0^2 = 0.0567$, $\hat{\rho} = -0.9047$ and $\hat{\xi} = 2.8055$. The Hull-White model SSPE of approximately 4 is a clear improvement over the Black-Scholes model SSPE of 240. Figure 3.17 gives point estimates and approximate pointwise 95% confidence intervals of the pricing errors along with the bid/ask spreads of the options. Clearly,

this model is appropriate for pricing these 29-day options.

To explore if this model is valid at another maturity, it is calibrated to the 50-day options in Table 3.8. Here, the algorithm terminated after only 8 iterations (see Figure 3.18) giving $\hat{\sigma}_0^2 = 0.0620$, $\hat{\rho} = -0.8062$ and $\hat{\xi} = 2.6672$ as the calibrating parameter values. Figure 3.19 gives point estimates and approximate pointwise 95% confidence intervals of the pricing errors along with the bid/ask spreads of the options. Again, this model is appropriate for pricing options at a single maturity.

There is a substantial difference between the 29-day and 50-day calibrated parameter values and it is interesting to examine if a single set of constant parameters can correctly price the options at both maturities. Thus, the model is calibrated using all of the options in Table 3.8. Figure 3.20 shows that the model is well-calibrated. That is, the average SSPE decreases to a neighbourhood and then randomly bounces around in this neighbourhood. The calibrated parameter values are $\hat{\sigma}_0^2 = 0.0594$, $\hat{\rho} = -0.8966$ and $\hat{\xi} = 2.6617$. Figures 3.21 and 3.22 show point estimates and approximate pointwise 95% confidence intervals of the pricing errors along with the bid/ask spreads of the options for the two maturities. While the model does a reasonable job of pricing in-the-money options at both maturities, it misprices at-the-money and out-of-the money options. This is evidence that the constant parameter zero-drift Hull-White model fails at pricing options across different maturities.

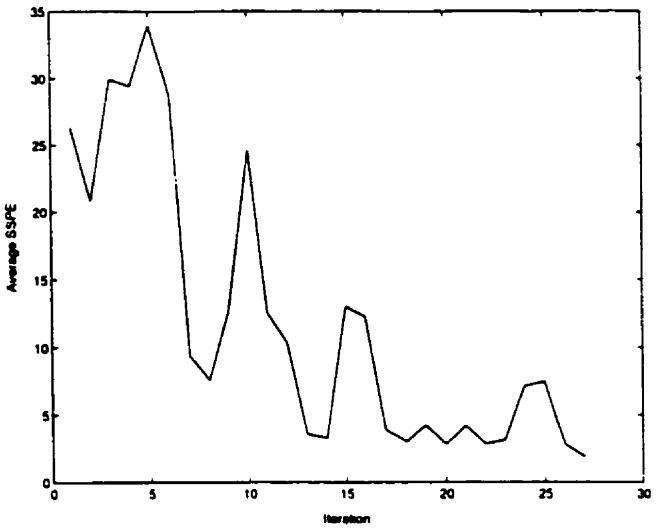


Figure 3.16: Average SSPE versus iteration number for the zero-drift Hull-White model calibration to 29-day S&P500 options.

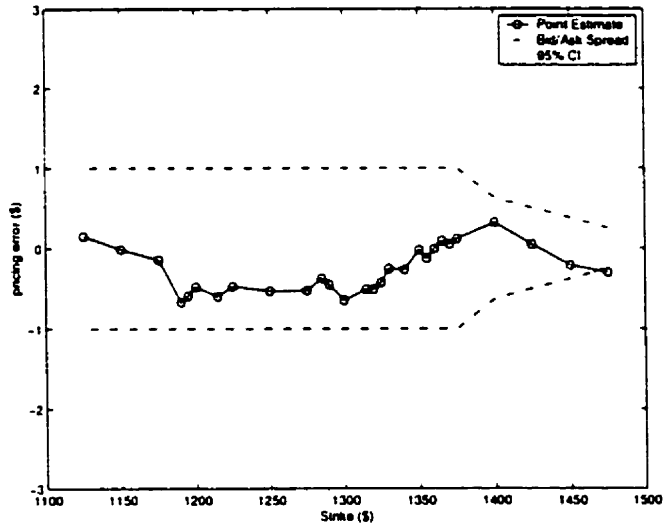


Figure 3.17: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of 29-day S&P500 call options using the zero-drift Hull-White model with $\hat{\sigma}_0^2 = 0.0567$, $\hat{\rho} = -0.9047$, and $\hat{\xi} = 2.8055$ based on 200,000 simulations (only 29-day options are used in the calibration).

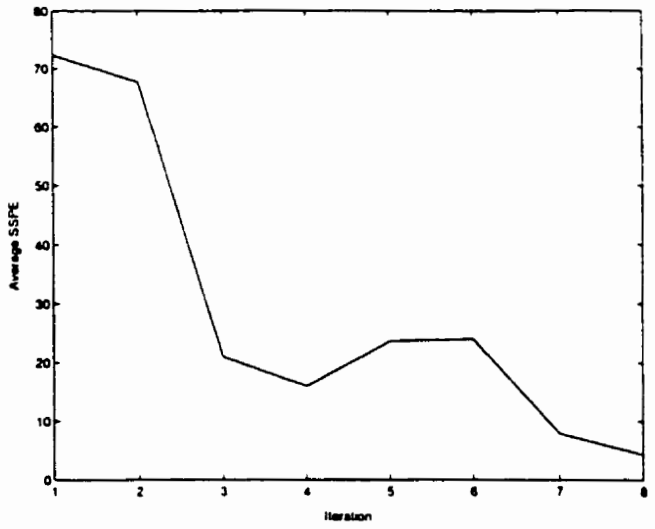


Figure 3.18: Average SSPE versus iteration number for the zero-drift Hull-White model calibration to 50-day S&P500 options.

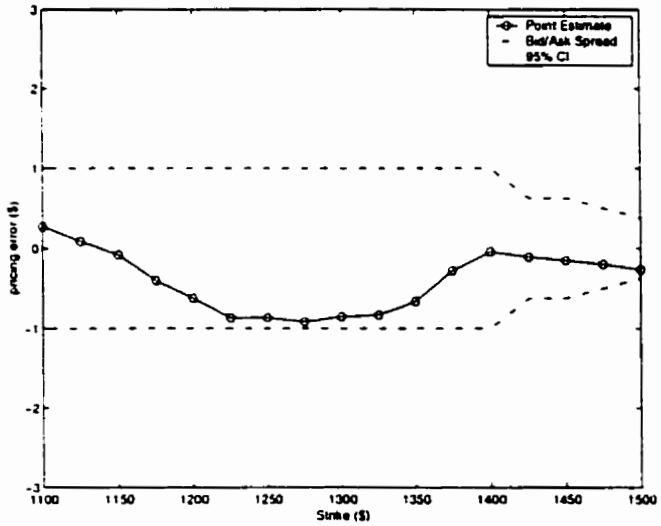


Figure 3.19: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of 50-day S&P500 call options using the zero-drift Hull-White model with $\hat{\sigma}_0^2 = 0.0620$, $\hat{\rho} = -0.8062$, and $\hat{\xi} = 2.6672$ based on 200,000 simulations (only 50-day options are used in the calibration).

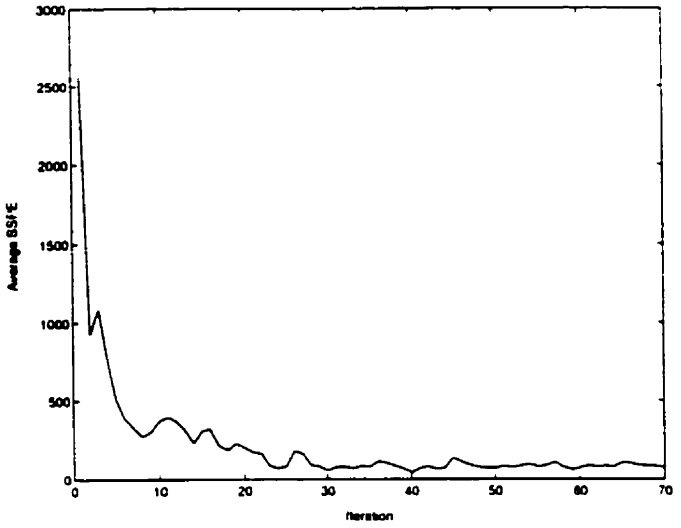


Figure 3.20: Average SSPE versus iteration number for the zero-drift Hull-White model calibration to 29-day and 50-day S&P500 options.

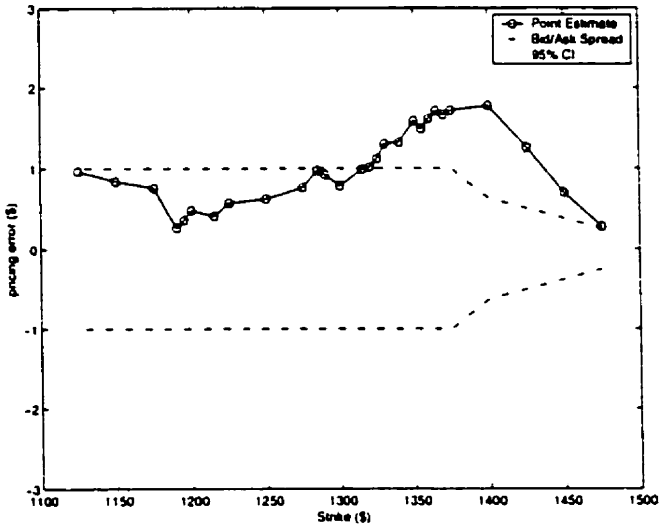


Figure 3.21: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of 29-day S&P500 call options using the zero-drift Hull-White model with $\hat{\sigma}_0^2 = 0.0594$, $\hat{\rho} = -0.8966$, and $\hat{\xi} = 2.6617$ based on 200,000 simulations (29-day and 50-day options are used in the calibration).

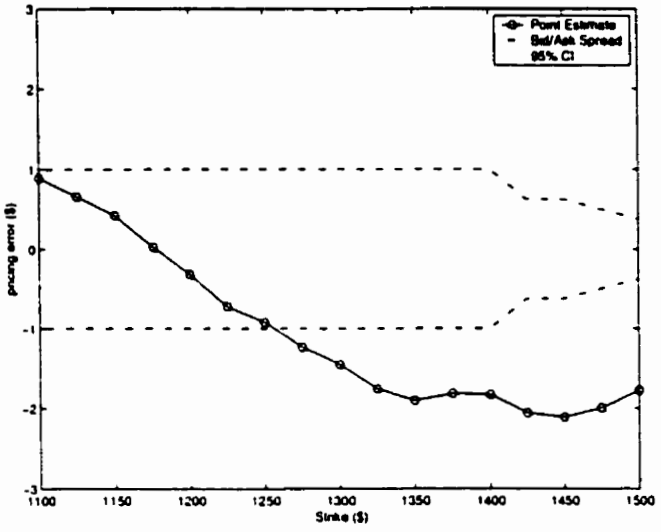


Figure 3.22: Point estimates and approximate pointwise 95% confidence intervals of the pricing error of 50-day S&P500 call options using the zero-drift Hull-White model with $\hat{\sigma}_0^2 = 0.0594$, $\hat{\rho} = -0.8966$, and $\hat{\xi} = 2.6617$ based on 200,000 simulations (29-day and 50-day options are used in the calibration).

## 3.5 Conclusion

The calibration method proposed here is successful at calibrating simulated models to option prices. It handles options from a range of moneyness and also allows for correlation between the stock and the variance processes as in the case of the Hull-White model. Experimental design methods are used to facilitate a polynomial approximation to the pricing error with a small number of experimental runs.

If a calibrated model matches market option prices at a single maturity it is considered an adequate model for these options (we have shown the Black-Scholes is inadequate at a single maturity). When a number of different maturities are considered, models with constant parameters typically fail. Model failure can be circumvented by using time and/or strike price dependent parameters (e.g., Dupire (1994), Derman and Kani (1994), Avellaneda et al (1997) and Andersen and Andreasen (1999)). We find that the calibrated Hull-White model is consistent with market prices at a single maturity but fails to allow for different maturities with constant parameters. This is clear evidence against the constant parameter zero-drift Hull-White model as the underlying risk-neutral data generating process. This failure across maturities is similarly observed for other models (e.g., Carr et al (2000) and Aït-Sahalia and Lo (1998)).

# Chapter 4

# Relative Entropy, Distortion and Risk

## 4.1 Introduction

As we have seen in previous chapters, relative entropy provides a method of deforming one probability distribution to another. Specifically, let $X$ be a random variable with cumulative distribution function (cdf) $F$ and $\mathcal{M}$ be a set of moment constraints that we want $X$ to satisfy under the new distribution. Using relative entropy as a measure of distance, the minimum relative entropy distribution (MRED) $\tilde{F}$ is the measure that is closest to $F$, satisfies the moment constraints $\mathcal{M}$ and is absolutely continuous with respect to $F$. The MRED $\tilde{F}$ is simply a reweighting of the measure $F$.

Another way to reweight a probability measure is to apply a suitable distortion function $g$ on the decumulative distribution function (ddf) $S$. The function $S^* =$

$g \circ S$ is a ddf and defines a distorted probability distribution (DPD) provided $g$ is a distortion function (Definition 4.3).

At first glance, the distortion approach of applying an operator on the ddf appears different than the relative entropy method of imposing certain moment constraints on the reweighted distribution. However, according to the Entropy Optimisation Postulate, the DPD can also be obtained as the solution to an entropy optimisation problem.

*Remark 4.1 (Entropy Optimisation Postulate).* Every probability distribution, theoretical or observed, is an entropy optimisation distribution, i.e., it can be obtained by maximising an appropriate entropy measure or by minimising a cross-entropy measure with respect to an appropriate a priori distribution, subject to its satisfying appropriate constraints (Kapur and Kesavan (1992) p.297).

We establish the connection between MRED's and DPD's and also show the relation between the set of moment constraints $\mathcal{M}$ imposed on the distribution and the distortion function $g$. This relation provides a method for constructing distortion functions through $\mathcal{M}$ as well as some intuition about the effect of a given distortion function on the original distribution. In Section 4.2.4 some examples show the equivalence of certain moment constraints with particular distortion functions.

Following this we discuss the application of DPD's and MRED's to measuring risks. Using the Choquet integral (Definition 4.6), we show how coherent risk measures are easily constructed through distortion. The link to relative entropy provides further insight into the properties of the resulting risk measure. Further-

more relative entropy and distortion provide a means for testing the coherence of a given risk measure. The final section discusses the pricing of financial assets using the Choquet integral and a distortion function suggested by Wang (1999). As seen in Example E8, the DPD generated by this distortion function does not correspond to an MRED with a constraint on the mean of the underlying asset. We develop a generalised measure of relative entropy that, when minimised subject to a constraint on the mean of the underlying asset, results in the same distribution as that produced through distortion. This shows that the distorted measure is consistent with typical pricing measures as it fixes the mean of the underlying asset.

In this chapter, we use the term density to refer to probability density functions (pdf) and probability functions (pf). As such, integrals and derivatives should be interpreted as sums and slopes where appropriate.

## 4.2 Relative Entropy and Distortion

### 4.2.1 Relative Entropy

Relative entropy, introduced in Section 1.4.2, provides a notion of distance from one probability distribution to another. Here, we restate its definition and the results showing how it is used as the objective function in a convex optimisation problem to determine MRED's.

**Definition 4.1 (Relative Entropy).** Suppose $f$ and $\tilde{f}$ are two pdf's. The *rela-*

*tive entropy*, $H(\tilde{f}|f)$, of $\tilde{f}$ with respect to $f$ is defined as

$$H(\tilde{f}|f) = \begin{cases} \int_\Omega \tilde{f}(x) \ln\left\{\frac{\tilde{f}(x)}{f(x)}\right\} dx & \text{if} \quad \tilde{f} \ll f \\ +\infty & \text{otherwise.} \end{cases} \tag{4.1}$$

The likelihood ratio $\frac{\tilde{f}}{f}$ is a Radon-Nikodym derivative of $\tilde{f}$ with respect to $f$.

**Definition 4.2 (MRED).** A *minimum relative entropy distribution* with pdf $f^*$ solves a convex optimisation problem. Specifically, given a random variable $X$ with *prior* pdf $f$, $f^*$ solves,

$$\begin{aligned} \min_{\tilde{f}} H(\tilde{f}|f) &= \min_{\tilde{f}} \left\{ \int_\Omega \tilde{f}(x) \ln\left\{\frac{\tilde{f}(x)}{f(x)}\right\} dx \right\} \qquad (4.2) \\ \text{subject to} \quad \int_\Omega \tilde{f}(x)dx &= 1, \\ E_{\tilde{f}}[G_j(X)] &= C_j, \quad \text{for} \quad 1 \le j \le N, \\ \tilde{f} &\ll f. \end{aligned}$$

where $G_1, \ldots, G_N$ are given functions and $C_1, \ldots, C_N$ are given constants.

Note that the first constraint combined with the absolute continuity constraint force $f^*$ to be a valid pdf. This optimisation problem is easily solved by Lagrangian methods, yielding an explicit expression for the MRED.

**Theorem 4.1.** *If there is a measure satisfying the constraints in 4.2, then the unique solution, $f^*$, to the constrained minimum relative entropy optimisation prob-*

*lem has the form*

$$f^*(x) = f(x)\exp\left\{\sum_{j=1}^{N}\lambda_j G_j(x) - \psi(\lambda)\right\}, \tag{4.3}$$

*where*

$$\psi(\lambda) = \ln\left(\int_\Omega f(x)\exp\left\{\sum_{j=1}^{N}\lambda_j G_j(x)\right\}dx\right),$$

*and* $\lambda = (\lambda_1, \ldots, \lambda_N)'$ *is determined uniquely by the moment constraints, that is*

$$\frac{\partial\psi}{\partial\lambda_j} = C_j, \qquad \textit{for } j = 1, \ldots, N.$$

### 4.2.2 Distortion

Distorted probability distributions are special cases of the more general theory of monotone set functions and non-additive measures. The reader is referred to the book by Denneberg (1994) for a careful treatment of the more general theory. Here, we do not use this level of abstraction and we caution the reader that the usual suspects of measure theory such as set functions and $\sigma$-algebras lurk in the background. We begin by defining a distortion function and a DPD.

**Definition 4.3 (Distortion Function).** A *distortion function*, $g$, is any non-decreasing function such that $g : [0,1] \mapsto [0,1]$ with $g(0) = 0$ and $g(1) = 1$.

**Definition 4.4 (DPD).** Let $X$ be a random variable with decumulative distribution function (ddf) $S$ and let $g$ be a distortion function. A *distorted probability*

*distribution*, $S^*$, is defined by

$$S^*(x) = g[S(x)]. \tag{4.4}$$

We use the ddf rather than the cdf as it is compatible with the Choquet integral (Definition 4.6). The dual distortion function is defined below and will be referred to in future sections.

**Definition 4.5 (Dual Distortion Function).** Given a distortion function $g$, the *dual distortion function of* $g$ is given by

$$\bar{g}(u) = 1 - g(1 - u). \tag{4.5}$$

For differentiable distortion functions, Equation 4.4 provides a way to compute the distorted probability density function (dpdf). Namely, when $X$ is continuous

$$f^*(x) = -\frac{d}{dx}S^*(x) = -\frac{d}{dx}g(S(x)) = f(x)g'(S(x)), \tag{4.6}$$

where $f$ is the original pdf of $X$. Furthermore, the DPD derived using the dual distortion function has pdf

$$f^*(x) = -\frac{d}{dx}\bar{g}\,(S^*(x)) = \frac{d}{dx}g(F(x)) = f(x)g'(F(x)). \tag{4.7}$$

This is exactly the density one obtains by using $F$ instead of $S$ in the distortion function $g$. The dpdf in the discrete case has a similar form, with the derivative replaced by a slope.

### 4.2.3 Link Between Relative Entropy and Distortion

Equations 4.6 and 4.3 reveal an obvious link between DPD's and MRED's. The theorem below formalises this relationship.

**Theorem 4.2 (DPD's and MRED's).** *Let $X$ be a random variable with pdf $f$ and ddf $S$.*

1. *If $S^*$ is the MRED satisfying a set of moment constraints $\mathcal{M}$, then $S^*$ is the DPD corresponding to some distortion function $g_{\mathcal{M}}$.*

2. *If $S^*$ is the DPD for a differentiable distortion function $g$, then $S^*$ is the MRED satisfying some set of moment constraints $\mathcal{M}_g$.*

Proof (continuous case):

1. Suppose that $S^*$ is an MRED, then by Theorem 4.1

$$f^*(x) = f(x)\exp\left(\sum_{j=1}^{N}\lambda_j G_j(x) - \psi(\lambda)\right). \tag{4.8}$$

Thus, by Equation 4.6

$$g'_{\mathcal{M}}(S(x)) = \exp\left(\sum_{j=1}^{N}\lambda_j G_j(x) - \psi(\lambda)\right). \tag{4.9}$$

The change of variable $u = S(x)$ and integration give

$$g_{\mathcal{M}}(u) = \int_0^u \exp\left\{\sum_{j=1}^{N}\lambda_j G_j\left(S^{-1}(v)\right) - \psi(\lambda)\right\} dv. \tag{4.10}$$

The condition $g_{\mathcal{M}}(0) = 0$ is obviously satisfied and Equation 4.8 ensures that $g_{\mathcal{M}}(1) = 1$ is satisfied.

2. Equation 4.6 shows that the dpdf is

$$f^*(x) = f(x)g'(S(x)) \tag{4.11}$$

and let us write $g'$ in the form

$$g'(S(x)) = \exp\left(\sum_{j=1}^{N} \lambda_j h_j(S(x)) - \psi(\lambda)\right) \tag{4.12}$$

for some functions $\psi$, $h_j$ and constants $\lambda_j$ determined by $\frac{\partial \psi}{\partial \lambda_j} = C_j$ for some constants $C_j$, $j = 1, \ldots, N$. By Theorem 4.1, $S^*$ is the MRED that solves

$$\begin{aligned} \min_{\tilde{f}} H(\tilde{f}|f) &= \min_{\tilde{f}} \left\{ \int_{\Omega} \tilde{f}(x) \ln \left\{ \frac{\tilde{f}(x)}{f(x)} \right\} dx \right\} \\ \text{subject to} \quad \int_{\Omega} \tilde{f}(x)dx &= 1, \\ E_{\tilde{f}}[h_j(S(X))] &= C_j, \quad \text{for} \quad 1 \leq j \leq N, \\ \tilde{f} &\ll f. \end{aligned} \tag{4.13}$$

A consequence of Theorem 4.2 is that distortion functions are implicitly defined through the formulation of a minimum relative entropy problem. The implied

distortion is given by

$$g(S(x)) = \int_0^{S(x)} \exp\left(\sum_{j=1}^{N} \lambda_j G_j\left(S^{-1}(u)\right) - \psi(\lambda)\right) du. \tag{4.14}$$

For the applications studied later in this chapter, it is of interest to know when a distortion function is concave or convex. The following theorem gives a necessary and sufficient condition on the moment constraints to ensure the concavity or convexity of the corresponding distortion function.

**Theorem 4.3.** *Suppose that a distortion function $g$ is defined by the moment constraints in a relative entropy optimisation problem. Thus $g$ is of the form given in Equation 4.14 and define the functions $h_j = G_j \circ S^{-1}$ for $j = 1, \ldots, N$. Then*

1. *$g$ is concave if and only if $\sum_{i=1}^{N} \lambda_i h_i'(u) \leq 0$ for all $u \in [0,1]$.*

2. *$g$ is convex if and only if $\sum_{i=1}^{N} \lambda_i h_i'(u) \geq 0$ for all $u \in [0,1]$.*

Proof (continuous case):

Differentiating $g'$ we get

$$g''(u) = \exp\left(\sum_{i=1}^{N} \lambda_i h_i(u) - \psi(\lambda)\right) \sum_{i=1}^{N} \lambda_i h_i'(u). \tag{4.15}$$

Since $g$ is a distortion function it is non-decreasing, hence $g' \geq 0$. Furthermore, since $g$ is a nonnegative twice differentiable function with $g' \geq 0$, we have

(a) $g$ is concave $\Leftrightarrow$ $g'' \leq 0$ $\Leftrightarrow \sum_{i=1}^{N} \lambda_i h_i'(u) \leq 0$ for all $u \in [0,1]$.

(b) $g$ is convex $\Leftrightarrow$ $g'' \geq 0$ $\Leftrightarrow \sum_{i=1}^{N} \lambda_i h_i'(u) \geq 0$ for all $u \in [0,1]$.

One can also construct more complicated distortion functions through mixing and composing existing ones. These operations preserve the property of concavity or convexity as stated in the following lemma (Wang 1996).

**Lemma 4.4.** *Let $g_i$, $i = 1, \ldots, n$ be concave (convex) distortion functions.*

1. *For $p_i \geq 0$ with $\sum_{i=1}^{n} p_i = 1$, the function $g = \sum_{i=1}^{n} p_i g_i$ is a concave (convex) distortion function.*

2. *The function $g = g_2 \circ g_1$ is a concave (convex) distortion function.*

Proof: The proof of this lemma is trivial.

### 4.2.4 Examples

In this section we provide a number of examples illustrating the distortion functions that certain moment constraints define. Table 4.1 summarises the examples and we discuss each one individually. Throughout this section, let $X$ be a continuous random variable with $f$, $F$ and $S$ its pdf, cdf and ddf respectively.

#### Example E1

The MRED is the conditional distribution of $X$ given that $X \in (F^{-1}(\alpha), F^{-1}(\beta)]$. This can be seen using the connection between DPD's and MRED's. The derivative

| | Constraint | Resulting Distortion |
|---|---|---|
| E1 | $P_{\tilde{f}}[1-\beta \le S(X) < 1-\alpha] = 1$<br>$0 \le \alpha < \beta \le 1$ | $g(u) = \min\left(\frac{(u-(1-\beta))^+}{\beta-\alpha}, 1\right)$ |
| E2 | $P_{\tilde{f}}[0 \le S(X) < \alpha] = \alpha(1+r)$<br>$0 < \alpha \le 1, \quad -1 \le r \le \frac{1}{\alpha} - 1$ | $g(u) = \min\left((1+r)u, \frac{\alpha r + (1-\alpha(1+r))u}{1-\alpha}\right)$ |
| E3 | $E_{\tilde{f}}[S(X)] = C$ | $g(u) = \frac{1-e^{-\alpha u}}{1-e^{-\alpha}}, \ \alpha > 0$ |
| E4 | $\begin{cases} E_{\tilde{f}}[-\ln(S(X))] = C_1 \\ E_{\tilde{f}}[-\ln(F(X))] = C_2 \end{cases}$<br>or $\begin{cases} E_{\tilde{f}}[\Theta(X)] = C_1 \\ E_{\tilde{f}}[\bar{\Theta}(X)] = C_2 \end{cases}$ | $g(u) = \int_0^u \frac{t^{a-1}(1-t)^{b-1}}{\beta(a,b)} dt$<br>$a > 0, \quad b > 0$ |
| E5 | $E_{\tilde{f}}[-\ln(S(X))] = C$<br>or $E_{\tilde{f}}[\Theta(X)] = C$ | $g(u) = u^a$<br>$a > 0$ |
| E6 | $E_{\tilde{f}}[-\ln(F(X))] = C$<br>or $E_{\tilde{f}}[\bar{\Theta}(X)] = C$ | $g(u) = 1 - (1-u)^b$<br>$b > 0$ |
| E7 | $E_{\tilde{f}}[\ln(a + bS(X))] = C$<br>$a > 0, \quad (a+b) > 0$ | $g(u) = \begin{cases} \frac{\ln(a+bu) - \ln a}{\ln(a+b) - \ln a}, & \text{if } \gamma = 0 \\ \frac{(a+bu)^\gamma - a^\gamma}{(a+b)^\gamma - a^\gamma}, & \text{if } \gamma > 0 \end{cases}$ |
| E8 | $E_{\tilde{f}}[\Phi^{-1}(S(X))] = C$<br>$\Phi$ is the $N(0,1)$ cdf | $g(u) = \Phi[\Phi^{-1}(u) + \alpha]$ |
| E9 | $E_{\tilde{f}}[h_1(X)] = C_1$<br>and $E_{\tilde{f}}[h_2(X)] = C_2$ | see Equation 4.14 |

Table 4.1: Moment constraints and resulting distortion functions.

of $g$ is

$$\begin{aligned} g'(u) &= \frac{I_{[1-\beta \le u < 1-\alpha]}}{\beta - \alpha} \\ &= \exp\left\{\ln\left(\frac{I_{[1-\beta \le u < 1-\alpha]}}{\beta - \alpha}\right)\right\} \\ &= \exp\left\{\lambda h(u) - \psi(\lambda)\right\}, \end{aligned} \tag{4.16}$$

where $\lambda = 1$, $h(u) = \ln\left(\frac{I_{[1-\beta \le u < 1-\alpha]}}{\beta - \alpha}\right)$ and $\psi(\lambda) = 0$. Then, Theorems 4.2 and 4.1 together imply that $C = \psi'(\lambda) = 0$ and the moment constraint is

$$E_{\tilde{f}}\left[\ln\left(\frac{I_{[1-\beta \le S(X) < 1-\alpha]}}{\beta - \alpha}\right)\right] = -\ln(\beta - \alpha), \tag{4.17}$$

which is true if and only if the indicator random variable equals one almost surely under $\tilde{f}$. Thus, the moment constraint is equivalent to

$$E_{\tilde{f}}\left[I_{[1-\beta \le S(X) < 1-\alpha]}\right] = 1. \tag{4.18}$$

showing indeed that the MRED is the conditional distribution of $X$ given that $X \in (F^{-1}(\alpha), F^{-1}(\beta)]$ (this is the reason for the term on the right side in Equation 4.17). If we take the limit as $\beta \downarrow \alpha$, we get a degenerate distribution at the quantile $\alpha$, that is $P_{\tilde{f}}[X = F^{-1}(\alpha)] = 1$. It is obvious to see that the distortion function $g$ is concave if and only if $\beta = 1$ and is convex if and only if $\alpha = 0$.

**Example E2**

This constraint leads to a piece-wise linear distortion function. The MRED is the conditional distribution of $X$ given that $P_f[F^{-1}(1-\alpha) < X \leq F^{-1}(1)] = \alpha(1+r)$. Another interpretation is that the $(1-\alpha)$ quantile, $F^{-1}(1-\alpha)$, of the original distribution becomes the $(1-\alpha(r+1))$ quantile of the MRED. The connection between the MRED and the DPD is determined using an argument with an indicator random variable similar to that used in Example E1. This distortion function is concave if and only if $\alpha(1+r) \geq \alpha \Leftrightarrow r \geq 0$ and is convex if and only if $r \leq 0$.

**Example E3: Exponential Distortion**

This distortion function is just the cdf of an exponential random variable constrained to the unit interval. The easiest way to determine the parameter $\alpha$ in terms of $C$ is to differentiate the distortion function and use the link between DPD's and MRED's. Here we have

$$\begin{aligned} g'(u) &= \frac{\alpha e^{-\alpha u}}{1-e^{-\alpha}} \\ &= \exp\left\{-\alpha u + \ln\left(\frac{\alpha}{1-e^{-\alpha}}\right)\right\} \\ &= \exp\{\lambda h(u) - \psi(\lambda)\} \end{aligned} \tag{4.19}$$

where $h(u) = u$, $\lambda = -\alpha$ and $\psi(\lambda) = -\ln\left(\frac{-\lambda}{1-e^{\lambda}}\right)$. Theorems 4.2 and 4.1 together imply that

$$C = \psi'(\lambda) = -\frac{1}{\lambda} - \frac{\lambda e^{\lambda}}{1-e^{\lambda}} = \frac{1}{\alpha} + \frac{\alpha e^{-\alpha}}{1-e^{-\alpha}}. \tag{4.20}$$

Furthermore, since $h'(u) = 1$ we have that $\lambda h'(u) = -\alpha$. Thus Theorem 4.3 implies that $g$ is concave if and only if $\alpha > 0$.

### Example E4: Beta Distortion

The distortion function is the incomplete beta function (the cdf of a $\beta(a,b)$ random variable) which is defined as

$$g(u) = \int_0^u \frac{1}{\beta(a,b)} t^{a-1}(1-t)^{b-1}dt, \tag{4.21}$$

where

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1}dt, \tag{4.22}$$

$\Gamma(a) = \int_0^{\infty} t^{a-1}e^{-t}dt$ is the gamma function and the parameters $a$ and $b$ are both positive. This transform includes the transforms in Examples E5 and E6 as special cases when $b = 1$ and $a = 1$ respectively.

As in the above example, we use the link between MRED's and DPD's to

determine the parameters $(a, b)$ in terms of $(C_1, C_2)$. The derivative of $g$ is

$$\begin{aligned} g'(u) &= \frac{1}{\beta(a,b)} u^{a-1}(1-u)^{b-1} \\ &= \exp\{(a-1)\ln(u) + (b-1)\ln(1-u) - \ln(\beta(a,b))\} \\ &= \exp\{\lambda_1 h_1(u) + \lambda_2 h_2(u) - \psi(\lambda_1, \lambda_2)\}, \end{aligned} \tag{4.23}$$

where $h_1(u) = -\ln(u)$, $h_2(u) = -\ln(1-u)$, $\lambda_1 = 1-a$, $\lambda_2 = 1-b$ and $\psi(\lambda_1, \lambda_2) = \ln(\beta(1-\lambda_1, 1-\lambda_2))$. Then, Theorems 4.2 and 4.1 together imply that

$$C_1 = \frac{\partial \psi}{\partial \lambda_1} = \frac{\Gamma'(1-\lambda_1+1-\lambda_2)}{\Gamma(1-\lambda_1+1-\lambda_2)} - \frac{\Gamma'(1-\lambda_1)}{\Gamma(1-\lambda_1)} \tag{4.24}$$

and

$$C_2 = \frac{\partial \psi}{\partial \lambda_2} = \frac{\Gamma'(1-\lambda_1+1-\lambda_2)}{\Gamma(1-\lambda_1+1-\lambda_2)} - \frac{\Gamma'(1-\lambda_2)}{\Gamma(1-\lambda_2)}. \tag{4.25}$$

In terms of the parameters $a$ and $b$

$$C_1 = \frac{\Gamma'(a+b)}{\Gamma(a+b)} - \frac{\Gamma'(a)}{\Gamma(a)} \quad \text{and} \quad C_2 = \frac{\Gamma'(a+b)}{\Gamma(a+b)} - \frac{\Gamma'(b)}{\Gamma(b)}. \tag{4.26}$$

From Theorem 4.3, $g$ is concave if and only if

$$\lambda_1 h_1'(u) + \lambda_2 h_2'(u) = \frac{a-1}{u} + \frac{1-b}{1-u} \leq 0 \tag{4.27}$$

for all $u \in [0, 1]$. It is easy to show the above condition is true if and only if $a \leq 1$ and $b \geq 1$. Similarly $g$ is convex if and only if $a \geq 1$ and $b \leq 1$.

By working with the moment constraints, more intuition about the differences between the MRED and the distribution of $X$ can be gleaned. First, let $f^*$, $F^*$ and $S^*$ be the pdf, cdf and ddf of the MRED respectively. Then under $F^*$, the random variables $S^*(X)$ and $F^*(X)$ are $U[0,1]$ random variables. Furthermore, $-\ln(S^*(X))$ and $-\ln(F^*(X))$ are exponential random variables with mean 1 under $F^*$. Therefore, the MRED satisfies the moment constraints

$$\begin{aligned} E_{f^*}\left[-\ln\left(\frac{S(X)}{S^*(X)}\right)\right] &= C_1 + 1 \qquad (4.28) \\ \text{and} \quad E_{f^*}\left[-\ln\left(\frac{F(X)}{F^*(X)}\right)\right] &= C_2 + 1. \end{aligned}$$

For another interpretation, note that for any continuous nonnegative random variable $X$, the ddf has the representation

$$S(x) = \exp\left\{-\int_0^x \theta(t)dt\right\} \qquad (4.29)$$

where $\theta(t) = \frac{f(x)}{S(x)}$ is the usual (forward-time) hazard rate. Similarly, the cdf has the representation

$$F(x) = \exp\left\{-\int_x^\infty \bar{\theta}(t)dt\right\}, \qquad (4.30)$$

where $\bar{\theta}(t) = \frac{f(x)}{F(x)}$ is the reverse-time hazard rate. Using these representations for

$S$ and $F$, the MRED can equivalently be derived using the moment constraints

$$\begin{aligned} E_{f}[\Theta(X)] &= C_1 \\ \text{and} \quad E_{f}[\bar{\Theta}(X)] &= C_2, \end{aligned} \tag{4.31}$$

where $\Theta(x) = \int_0^x \theta(t)dt$ and $\bar{\Theta}(x) = \int_x^\infty \bar{\theta}(t)dt$ are the integrated forward and reverse-time hazard rates of the original distribution respectively.

Providing that we can interchange the order of integration, we obtain

$$\begin{aligned} E_{f^*}[\Theta(X)] &= \int_0^\infty \left( \int_0^x \frac{f(y)}{S(y)} dy \right) f^*(x)dx \\ &= \int_0^\infty \left( \int_y^\infty f^*(x)dx \right) \frac{1}{S(y)} f(y)dy \\ &= \int_0^\infty \frac{S^*(y)}{S(y)} f(y)dy \\ &= E\left[ \frac{S^*(X)}{S(X)} \right]. \end{aligned} \tag{4.32}$$

where the last expectation is under the original distribution. A similar exercise with the other moment constraint shows that the MRED satisfies

$$\begin{aligned} E\left[ \frac{S^*(X)}{S(X)} \right] &= C_1 \\ \text{and} \quad E\left[ \frac{F^*(X)}{F(X)} \right] &= C_2, \end{aligned} \tag{4.33}$$

where the expectations are taken under the original measure. Using the likelihood

ratio

$$\frac{f(x)}{f^*(x)} = \frac{1}{g'(S(x))} = \frac{\beta(a,b)}{S(x)^{a-1}F(x)^{b-1}}, \tag{4.34}$$

the expectations in the above constraints can be changed to expectations under the MRED. Specfically, the MRED also satisfies

$$\begin{aligned} E_{f^*}\left[\frac{S^*(X)}{S(X)^a F(X)^{b-1}}\right] &= \frac{C_1}{\beta(a,b)} \\ \text{and} \quad E_{f^*}\left[\frac{F^*(X)}{F(X)^b S(X)^{a-1}}\right] &= \frac{C_2}{\beta(a,b)}. \end{aligned} \tag{4.35}$$

**Example E5: Proportional Hazards Distortion**

As previously mentioned, this transform is a special case of the beta distortion obtained by setting $b = 1$. Although the analysis of the previous section applies, in this case there is a much simpler relationship between $C$ and $a$. The derivative of $g$ is

$$\begin{aligned} g'(u) &= au^{a-1} \\ &= \exp\{(a-1)\ln(u) + \ln(a)\} \\ &= \exp\{\lambda h(u) - \psi(\lambda)\}, \end{aligned} \tag{4.36}$$

where $h(u) = -\ln(u)$, $\lambda = 1 - a$, and $\psi(\lambda) = \ln\left(\frac{1}{1-\lambda}\right)$. Theorems 4.2 and 4.1 together imply that

$$C = \psi'(\lambda) = \lambda - 1 = -a. \tag{4.37}$$

Furthermore, we have $\lambda h'(u) = (a-1)\frac{1}{u}$. Thus by Theorem 4.3, $g$ is concave if and only if $a \leq 1$ and is convex if and only if $a \geq 1$. Therefore, $g$ is either concave or convex (or both if $a = 1$).

All of the different formulations of the moment constraints performed for the beta distortion apply. Using the form for the ddf in Equation 4.29 we see that

$$S^*(x) = \left(\exp\left\{-\int_0^x \theta(t)dt\right\}\right)^a = \exp\left\{-\int_0^x a\theta(t)dt\right\}. \tag{4.38}$$

Therefore the DPD has a hazard rate of $a\theta$, hence the name proportional hazards (PH) distortion.

### Example E6: Dual Power Distortion

The dual power distortion is a special case of the beta distortion obtained by setting $a = 1$. From Definition 4.5, we see that $g$ is the dual distortion function to the PH distortion, hence the name dual power distortion. As with the PH transform, it is

instructive to obtain the relationship between $C$ and $b$. The derivative of $g$ is

$$\begin{aligned} g'(u) &= b(1-u)^{b-1} \\ &= \exp\{(b-1)\ln(1-u) + \ln(b)\} \\ &= \exp\{\lambda h(u) - \psi(\lambda)\}, \end{aligned} \tag{4.39}$$

where $h(u) = -\ln(1-u)$, $\lambda = 1-b$, and $\psi(\lambda) = \ln\left(\frac{1}{1-\lambda}\right)$. Theorems 4.2 and 4.1 together imply that

$$C = \psi'(\lambda) = \lambda - 1 = -b. \tag{4.40}$$

Furthermore, we have $\lambda h'(u) = \frac{1-b}{1-u}$. Thus by Theorem 4.3, $g$ is concave if and only if $b \geq 1$ and is convex if and only if $b \leq 1$. Therefore, $g$ is either concave or convex (or both if $b = 1$). As with the PH transform, all of the different moment formulations from the beta distortion section are valid for the dual power distortion.

**Example E7**

This moment constraint implies both logarithmic and power distortion functions depending on the value of the constant $C$. As usual, we exploit the link between DPD's and MRED's to determine the parameter $C$.

**Logarithmic Distortion ($\gamma = 0$)**

The derivative of the distortion function is

$$\begin{aligned} g'(u) &= \frac{1}{\ln(a+b) - \ln(a)} \left( \frac{b}{a+bu} \right) \\ &= \exp\left\{ -\ln\left[ (a+bu) \left( \frac{b}{\ln\left(\frac{a+b}{a}\right)} \right)^{-1} \right] \right\} \\ &= \exp\left\{ \lambda h(u) - \psi(\lambda) \right\}, \end{aligned} \tag{4.41}$$

where $\psi(\lambda) = 0$, $\lambda = 1$ and

$$h(u) = -\ln\left[ (a+bu) \left( \frac{b}{\ln\left(\frac{a+b}{a}\right)} \right)^{-1} \right]. \tag{4.42}$$

Clearly, we have that $\psi'(\lambda) = 0$ which would lead one to think that $C = 0$. However, we can write $h$ as

$$h(u) = -\ln(a+bu) + \ln\left( \frac{b}{\ln\left(\frac{a+b}{a}\right)} \right), \tag{4.43}$$

which leads to the moment constraint

$$E_f\left[\ln(a+bS(X))\right] = \ln\left( \frac{\ln\left(\frac{a+b}{a}\right)}{b} \right) = C. \tag{4.44}$$

For conditions when $g$ is concave or convex, we examine $\lambda h'(u) = \frac{-b}{a+bu}$ and see that $g$ is concave if and only if $b \geq 0$ and is convex if and only if $b \leq 0$ (by Theorem 4.3). The reader is reminded of the other restrictions $a > 0$ and $(a+b) > 0$.

**Power Distortion ($\gamma \neq 0$)**

In the case of the power distortion, the derivative is

$$\begin{aligned} g'(u) &= \frac{(a+bu)^{\gamma-1}b\gamma}{(a+b)^{\gamma}-a^{\gamma}} \\ &= \exp\left\{(\gamma-1)\ln(a+bu)+\ln\left(\frac{b\gamma}{(a+b)^{\gamma}-a^{\gamma}}\right)\right\} \\ &= \exp\left\{\lambda h(u)-\psi(\lambda)\right\}, \end{aligned} \tag{4.45}$$

where $h(u) = \ln(a+bu)$, $\lambda = \gamma - 1$ and $\psi(\lambda) = \ln\left(\frac{(a+b)^{\lambda+1}-a^{\lambda+1}}{(\lambda+1)b}\right)$. Theorems 4.2 and 4.1 together imply that

$$C = \psi'(\lambda) = \frac{(a+b)^{\gamma}\ln(a+b)-a^{\gamma}\ln(a)}{(a+b)^{\gamma}-a^{\gamma}} - \frac{1}{\gamma}. \tag{4.46}$$

Furthermore, we have $\lambda h'(u) = (\gamma-1)\frac{b}{a+bu}$. Thus by Theorem 4.3, $g$ is concave if and only if either ($b \geq 0$ and $\gamma \leq 1$) or ($-a < b \leq 0$ and $\gamma \geq 1$). Similarly $g$ is convex if and only if either ($b \geq 0$ and $\gamma \geq 1$) or ($-a < b \leq 0$ and $\gamma \leq 1$). Of course, this transform is a square root transform for $\gamma = \frac{1}{2}$ and a quadratic transform for $\gamma = 2$.

**Example E8: Normal Distortion**

Through the inverse probability transform, the random variable $X$ is mapped to a normal random variable whose mean is then shifted by $\alpha$. It turns out that this distortion is a special case of the exponential family distortions discussed in Section 4.5. As usual, the derivative of $g$ helps to get the relation between $C$ and $\alpha$. In

this case, the derivative is

$$\begin{aligned} g'(u) &= \frac{\phi(\Phi^{-1}(u)+\alpha)}{\phi(\Phi^{-1}(u))} \\ &= \exp\left\{-\alpha\Phi^{-1}(u) - \frac{\alpha^2}{2}\right\} \\ &= \exp\left\{\lambda h(u) - \psi(\lambda)\right\}, \end{aligned} \tag{4.47}$$

where $\phi$ is the standard normal density function, $\lambda = -\alpha$, $h(u) = \Phi^{-1}(u)$ and $\psi(\lambda) = \frac{\lambda^2}{2}$. Theorems 4.2 and 4.1 together imply that

$$C = \psi'(\lambda) = \lambda = -\alpha. \tag{4.48}$$

Furthermore, we have $\lambda h'(u) = \frac{-\alpha}{\phi(\Phi^{-1}(u))}$. Thus by Theorem 4.3, $g$ is concave if and only if $\alpha \geq 0$ and is convex if and only if $\alpha \leq 0$. Also, the dual distortion function of $g$ is

$$\bar{g}(u) = 1 - g(1-u) = \Phi\left[\Phi^{-1}(u) - \alpha\right] \tag{4.49}$$

which is just the inverse of $g$, a consequence of the symmetry property of the normal distribution.

### Example E9

In a relative entropy problem it is sometimes more natural to think about imposing moment constraints on the random variable $X$ rather than on the ddf. Note that

we can write any function $h$ as

$$h(x) = h\left(S^{-1}(S(x))\right). \tag{4.50}$$

This explicitly shows that the moment constraints in this example can be written as

$$\begin{aligned} E_{\hat{f}}\left[\bar{h}_1(S(X))\right] &= C_1 \\ \text{and} \quad E_{\hat{f}}\left[\bar{h}_2(S(X))\right] &= C_2, \end{aligned} \tag{4.51}$$

where $\bar{h}_1 = h_1 \circ S^{-1}$ and $\bar{h}_2 = h_2 \circ S^{-1}$. The derivative of the distortion function is

$$g'(u) = \exp\left\{\lambda_1 h_1\left(S^{-1}(u)\right) + \lambda_2 h_2\left(S^{-1}(u)\right) - \psi(\lambda_1, \lambda_2)\right\}, \tag{4.52}$$

which, upon substituting $u = S(x)$ becomes

$$g'(S(x)) = \exp\left\{\lambda_1 h_1(x) + \lambda_2 h_2(x) - \psi(\lambda_1, \lambda_2)\right\}. \tag{4.53}$$

With $h_2 = 0$, this is the generalised Esscher transform given in Kamps (1998). That is, the distribution obtained through a generalised Esscher transform is exactly the MRED obtained by constraining the expected value of $h_1(X)$. If $h_1(u) = u$ in addition to $h_2 = 0$ the MRED is the distribution obtained by the usual Esscher transform.

From an asset pricing perspective, arbitrage considerations typically prescribe the mean of the underlying asset, which gives $h_1(u) = u$ and $h_2 = 0$. In addition to

the constraint on the mean, suppose we wish to impose a calibrating volatility on the underlying asset. This is accomplished with $h_1(u) = u$ and $h_2(u) = u^2$. These are exactly the constraints used in the option pricing models developed in Chapter 2.

## 4.3 Choquet Integral

In this section, we define the Choquet integral and list some of its well-known properties. Properties about the Choquet integral that are useful in the construction of insurance premium principles and risk measures with desirable characteristics are then provided. Since many insurance premium principles and risk measures have a Choquet integral representation, (Wang (1996), Wang et al (1997), Wirch and Hardy (1999)), these properties may also be used to verify if a given risk measure or premium principle has a certain characteristic. Desirable characteristics of risk measures and premium principles are briefly mentioned in passing here and are discussed more fully in the next section. Although this section focusses almost entirely on distortion functions, the reader is reminded of the equivalence of relative entropy and distortion.

**Definition 4.6 (Choquet Integral).** For any random variable $X$ with ddf $S$, the *Choquet integral* with respect to distortion function $g$ is defined by

$$H_g(X) = \int_{-\infty}^{0} \{g[S(x)] - 1\}dx + \int_{0}^{\infty} g[S(x)]dx \tag{4.54}$$

(Denneberg (1994)).

Note that $H_g(\cdot)$ is not to be confused with the notation for entropy and relative entropy.

**Lemma 4.5.** *For any random variable $X$ with ddf $S$ and any distortion function $g$, the Choquet integral is equivalent to the expected value of $X$ under the DPD with ddf $S^*$. That is,*

$$H_g(X) = E_{S^*}[X], \tag{4.55}$$

*where $E_{S^*}[\cdot]$ denotes expectation under the distorted measure.*

Proof:

Define $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. Then $X$ can be written as $X = X^+ - X^-$ and the expected value is

$$E_{S^*}[X] = E_{S^*}\left[X^+\right] - E_{S^*}\left[X^-\right]. \tag{4.56}$$

It is easily shown that

$$E_{S^*}\left[X^+\right] = \int_0^\infty S^*(x)dx \quad \text{and} \quad E_{S^*}\left[X^-\right] = \int_{-\infty}^0 (1 - S^*(x))\,dx. \tag{4.57}$$

Combining these expressions with Equation 4.56 we have

$$\begin{aligned} E_{S^*}[X] &= \int_{-\infty}^0 (S^*(x) - 1)\,dx + \int_0^\infty S^*(x)dx \\ &= \int_{-\infty}^0 (g(S(x)) - 1)\,dx + \int_0^\infty g(S(x))dx \\ &= H_g(X). \end{aligned} \tag{4.58}$$

**Definition 4.7 (Comonotonic Random Variables).** Two random variables $X$ and $Y$ are *comonotonic* if there exists another random variable $Z$ and increasing real-valued functions $u$ and $v$ such that

$$X = u(Z) \qquad Y = v(Z). \tag{4.59}$$

A class ($\mathcal{C}$) of random variables is called *comonotonic* if and only if each pair of random variables in $\mathcal{C}$ is comonotonic (Denneberg (1994)).

If $X$ and $Y$ represent risks and are comonotonic, this means that one is not a hedge for the other. We now list some well-known properties of the Choquet integral (Denneberg (1994) Chapters 5 and 6).

**Theorem 4.6 (Properties of $H_g$).** *For any distortion function $g$ and real-valued random variables $X, Y$, the following properties hold:*

1. *Monotonicity*

   $X \leq Y$ *implies* $H_g(X) \leq H_g(Y)$.

2. *Positive Homogeneity*

   $H_g(cX) = cH_g(X)$ *for* $c \geq 0$.

3. *If $X = c$ for any constant $c$, then $H_g(X) = c$*

4. *Translation Invariant*

   $H_g(X + c) = H_g(x) + c$ *for any constant $c$.*

5. *Comonotonic Additivity*

   *If $X, Y$ are comonotonic, then* $H_g(X + Y) = H_g(X) + H_g(Y)$.

6. *Subadditive for Concave g*

   *If g is concave then* $H_g(X+Y) \leq H_g(X) + H_g(Y)$.

7. *Superadditive for Convex g*

   *If g is convex then* $H_g(X+Y) \geq H_g(X) + H_g(Y)$.

8. *Asymmetry*

   $H_g(-X) = -H_{\bar{g}}(X)$ *where* $\bar{g}$ *is the dual distortion function of g.*

Note that the monotonicity property above ensures that

$$H_g(X) \leq \max(X) \tag{4.60}$$

for any distortion $g$ and any random variable $X$. This corresponds to the principle of non-excessive loading. That is, the price of a risk should not exceed the maximum possible loss. Subadditivity is another nice property for a risk measure, which simply states that there should be no incentive to split risks. The following proposition addresses the issue of subadditivity of the Choquet integral.

**Proposition 4.7.** *Let g be a twice continuously differentiable distortion function. Then, g is concave if and only if* $H_g$ *is subadditive.*

Proof:

1. Sufficiency is given by Theorem 4.6 and is true for all concave distortions (not necessarily differentiable).

2. A modified version of a proof from Wirch and Hardy (2000). Concavity as a necessary condition is proven via a contrapositive argument. Suppose that $g$

| | | $Y$ | | | |
|---|---|---|---|---|---|
| | | $0$ | $w+\frac{z}{2}$ | $x+z$ | $\Sigma$ |
| $X$ | $0$ | $1-c-z$ | $z$ | $0$ | $1-c$ |
| | $w+z$ | $z$ | $0$ | $c-z$ | $c$ |
| | $\Sigma$ | $1-c$ | $z$ | $c-z$ | $1$ |

Table 4.2: Joint distribution of $X, Y$.

is not concave. Then, since $g$ is twice continuously differentiable, there is at least one interval $(a,b) \subseteq [0,1]$ where $g$ is strictly convex. Let $c = \frac{a+b}{2}$ be the midpoint of the interval $(a,b)$. Then for any $z \in \left(0, \frac{b-a}{2}\right)$ the strict convexity of $g$ implies

$$g(c+z) - g(c) > g(c) - g(c-z). \tag{4.61}$$

Now, consider the random variables $X$ and $Y$ with the discrete joint distribution for any $w > 0$ and for any $z \in \left(0, \frac{b-a}{2}\right)$ given in Table 4.2. The Choquet integrals are

$$\begin{aligned} H_g(X) &= (w+z)g(c). \\ H_g(Y) &= \left(w+\frac{z}{2}\right)g(c) + \frac{z}{2}g(c-z) \quad \text{and} \\ H_g(X+Y) &= \left(w+\frac{z}{2}\right)g(c+z) + \frac{z}{2}g(c) + (w+z)g(c-z). \end{aligned}$$

Subtracting we see that

$$\begin{aligned} H_g(X+Y) - (H_g(X) + H_g(Y)) & \\ &= \left(w + \tfrac{z}{2}\right)\{[g(c+z) - g(c)] - [g(c) - g(c-z)]\} \qquad (4.62) \\ &> 0 \quad \text{from Equation 4.61.} \end{aligned}$$

Thus, for a twice continuously differentiable distortion function $g$, we have shown that if $g$ is not concave then there exist random variables $X$ and $Y$ where $H_g(X+Y) > H_g(X) + H_g(Y)$ (i.e., the Choquet integral is not subadditive). Therefore, by a contrapositive argument we have shown that $H_g(X+Y) \leq H_g(X) + H_g(Y)$ for all random variables $X$ and $Y$ implies that $g$ is concave.

For insurance and risk management applications of the Choquet integral, it is useful to know conditions on the distortion function that ensure the integral is bounded below by the expected value of the random variable (nonnegative loading).

**Proposition 4.8.** *For a distortion function $g$, $g(u) \geq u$ for all $u \in [0,1]$ if and only if $E[X] \leq H_g(X)$ for all random variables $X$.*

Proof:

1. Suppose $g(u) \geq u$ for all $u \in [0,1]$. Using the identity function as the distor-

tion function in Lemma 4.5 we have

$$\begin{aligned} E[X] &= \int_{-\infty}^{0} (S(x)-1)\,dx + \int_{0}^{\infty} S(x)dx \\ &\leq \int_{-\infty}^{0} (g(S(x))-1)\,dx + \int_{0}^{\infty} g(S(x))dx \\ &= H_g(X). \end{aligned} \tag{4.63}$$

where the inequality follows from the assumption that $g(u) \geq u$ for all $u \in [0,1]$.

2. A modified version of a proof from Wirch and Hardy (2000). We prove the other direction with a contrapositive argument. Suppose there exists $u_0 \in [0,1]$ such that $g(u_0) < u_0$ and define the random variable

$$X = \begin{cases} 0 & \text{w.p.} \quad 1-u_0 \\ w & \text{w.p.} \quad u_0 \end{cases} \tag{4.64}$$

for any $w > 0$. Then

$$H_g(X) = wg(u_0) < wu_0 = E[X]. \tag{4.65}$$

Therefore, by contrapositive we have shown that $H_g(X) \geq E[X]$ for all random variables implies that $g(u) \geq u$ for all $u \in [0,1]$.

For the purposes of constructing risk measures via the Choquet integral and distortion (or relative entropy), the following result says that concavity of the distortion function $g$ is sufficient to ensure that $H_g$ is subadditive and bounded below

by the mean.

**Corollary 4.9.** *If $g$ is a concave distortion function then $H_g$ is subadditive and* $E[X] \leq H_g(X)$.

Proof:

Theorem 4.6 gives the subadditivity property. By noting that concavity implies $g(u) \geq u$ for all $u \in [0,1]$, Proposition 4.8 gives $E[X] \leq H_g(X)$.

In Theorem 4.2 we show the equivalence of DPD's and MRED's. The following corollary ties together i) Choquet integrals that are subadditive and bounded below by the mean, ii) concave distortion functions, and iii) conditions on the moment constraints in the corresponding relative entropy optimisation problem. This may be used to test whether a given risk measure with a Choquet integral representation is subadditive and bounded below by the expected value.

**Corollary 4.10.** *For all random variables and a twice continuously differentiable distortion function $g$ with*

$$g'(u) = \exp\left(\sum_{i=1}^{N} \lambda_i h_i(u) - \psi(\lambda)\right), \tag{4.66}$$

*where $h_i$ are given functions and $\lambda = (\lambda_1, \ldots, \lambda_N)'$ and $\psi$ satisfy $\frac{\partial \psi}{\partial \lambda_i} = C_i$ for given numbers $C_i$, $i = 1, \ldots, N$, the following are equivalent:*

1. *$H_g$ is subadditive and $E[X] \leq H_g(X)$.*

2. *$g$ is a concave distortion function.*

3. $\sum_{i=1}^{N} \lambda_i h_i'(u) \leq 0$.

Proof:

Proposition 4.7 shows that $1 \Rightarrow 2$.

Corollary 4.9 shows that $1 \Leftarrow 2$.

Theorem 4.3 shows that $2 \Leftrightarrow 3$ (continuous case).

## 4.4 Measuring Risks

Risk measures are used extensively in finance and insurance. Prices of risks such as an insurance premium are determined by measuring the risks associated with the product. Exchanges and clearing houses determine margin requirements by measuring the riskiness of an investor's portfolio. Risk measures are also used to set capital requirements to help ensure the solvency of a company. These requirements may be imposed by a regulator or by a company's internal risk managment protocol. Companies with many lines of business use risk measures to rate the performance of each business line through its risk-adjusted return on capital (RAROC). The need for good risk measures is quite apparent.

Recently, the Choquet integral has been revealed as an important tool in the measuring and pricing of financial and insurance risks. Chateauneuf et al (1996) use it to explain apparent discrepancies in observed market prices, such as violation of put-call parity, due to friction in the market caused by the bid/ask spread. The Choquet integral with normal distortion of Example E8 has been proposed

to price both financial and insurance risks (Wang (1999)). Many recent papers in the insurance literature illustrate the use of the Choquet integral as a premium principle. Some examples are Wang (1995, 1996), Wang and Young (1998), Kamps (1998) and Wang and Dhaene (1998).

By taking an axiomatic approach to insurance prices, Wang et al (1997) show that any market premium functional satisfying the prescribed axioms has a Choquet integral representation. Artzner et al (1999) use a similar axiomatic approach for general risk measures. We outline the axioms of Artzner et al in the following section and show the Choquet integral is useful not only in constructing good risk measures, but also in testing the properties of a given risk measure (Wirch (1999)). The link between relative entropy and distortion gives further insight about risk measures with a Choquet integral representation.

### 4.4.1 Coherent Risk Measures

The definitions and axioms presented here are slightly modified versions of the definitions and axioms given by Artzner et al (1999).

**Definition 4.8 (Risk).** A *risk*, $X$, is a random variable representing the future net loss of an investor at some particular time in the future. If $X < 0$ this represents a gain and $X \geq 0$ represents a loss.

**Definition 4.9 (Risk Measure).** A *risk measure*, $\rho$, is a functional mapping $X$ to the real line, that is $\rho : X \mapsto \Re$.

For simplicity, we assume that the risk-free rate of interest is zero. Artzner et al (1999) give the following axioms that are desirable characteristics of a risk measure:

1. **Monotonicity**

   If $P(X \leq Y) = 1$ then $\rho(X) \leq \rho(Y)$.

2. **Positive Homogeneity**

   For all $\lambda \geq 0$, $\rho(\lambda X) = \lambda\rho(X)$.

3. **Translation Invariance**

   For all risks and real numbers $\alpha$, $\rho(X + \alpha) = \rho(X) + \alpha$.

4. **Subadditivity**

   For all risks $X, Y$, $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

5. **Relevance**

   For all nonnegative risks with $P(X > 0) > 0$, $\rho(X) > 0$.

The monotonicity axiom says that if risk $Y$ is always greater than risk $X$, then the risk measure for $Y$ should also be greater than the risk measure for $X$. It ensures the property of non-excessive loading is satisfied, ruling out the standard deviation principle. Positive homogeneity reflects that the size ($\lambda$) of a position taken on risk $X$ increases the risk measure associated with $X$ by a factor of $\lambda$. Translation invariance means that adding (subtracting) the sure amount $\alpha$ to the risk $X$ increases (decreases) the risk measure by $\alpha$. That is, if $\alpha$ units of the risk-free asset is removed from a portfolio, then the risk measure should increase by the same amount. Subadditivity can be restated as "a merger does not create extra risk", or "there is no incentive to split risks". The relevance axiom is a necessary, but not sufficient, condition to prevent concentration of risks to remain undetected.

**Definition 4.10 (Coherent Risk Measure).** A risk measure is *coherent* if it satisfies the four axioms of monotonicity, positive homogeneity, translation invariance and subadditivity.

From Proposition 4.1 in Artzner et al (1999), we see that if $\rho$ is a coherent risk measure, then it is bounded below by the mean net loss, that is

$$\rho(X) \geq E[X] \tag{4.67}$$

for all risks $X$. We will refer to this as the nonnegative loading property to be consistent with insurance premium terminology.

From Theorem 4.6 we see that for any distortion function, the Choquet integral satisfies the axioms of monotonicity, positive homogeneity and translation invariance. The other results in Section 4.3 address necessary and sufficient conditions on the distortion function for the Choquet integral to be subadditive and satisfy the nonnegative loading property. As such, the Choquet integral is a natural tool for constructing coherent risk measures by specifying an appropriate distortion function. Furthermore, if a given risk measure has a Choquet integral representation, the results from Section 4.3 allow one to verify if the risk measure is coherent.

**Definition 4.11 (Distorted Risk Measure).** A risk measure is a *distorted risk measure* if it has a Choquet integral representation. The notation $\rho_g$ is used to denote the distorted risk measure with distortion function $g$.

Commonly used risk measures such as Value-at-Risk (VaR) and Tail-VaR have a Choquet integral representation as discussed in Section 4.4.2, hence they are exam-

ples of distorted risk measures (Wirch and Hardy (1999)). Corollary 4.9 provides sufficient conditions on the distortion function to ensure that a distorted risk measure is coherent. This is useful for constructing coherent risk measures. We restate the result in terms of coherence of the distorted risk measure.

**Corollary 4.11.** *If $g$ is a concave distortion function then the distorted risk measure $\rho_g(X)$ is coherent.*

Typically risk measures are used to compute capital requirements or reserves to protect a company against ruin. A risk measure that considers only the loss part of the distribution is conservative as potential losses are not offset by potential gains. As stated in Artzner et al (1999) Remark 4.5, actuaries have been considering only the loss part of the distribution since the 1800's. The condition $g(u) \geq \min\left(1, \frac{u}{S(0)}\right)$ for all $u \in [0,1]$ is sufficient to ensure the relevance axiom is satisfied as only the loss part of the distribution is used to calculate the distorted risk measure. However, it is quite strong and can probably be relaxed somewhat for particular distributions. Finally, note that distorted risk measures always satisfy the relevance axiom for nonnegative random variables.

As previously stated, the properties of the Choquet integral can be used to test the coherence of a given distorted risk measure. The following corollary provides a test for certain distortion functions and also reminds us of the link to relative entropy (compare with Corollary 4.10).

**Corollary 4.12.** *For all random variables and a twice continously differentiable*

*distortion function $g$ with*

$$g'(u) = \exp\left(\sum_{i=1}^{N} \lambda_i h_i(u) - \psi(\lambda)\right), \tag{4.68}$$

*where $h_i$ are given functions and $\lambda = (\lambda_1, \ldots, \lambda_N)'$ and $\psi$ satisfy $\frac{\partial \psi}{\partial \lambda_i} = C_i$ for given numbers $C_i$, $i = 1, \ldots, N$, the following are equivalent:*

1. *the distorted risk measure $\rho_g(X)$ is coherent.*

2. *$g$ is a concave distortion function.*

3. *$\sum_{i=1}^{N} \lambda_i h_i'(u) \leq 0$.*

### Risk Preferences Through Distortion

For all random variables, the distorted risk measure is $\rho_g(X) = E_{S^*}[X]$ where the expectation is taken with respect to the DPD (equivalently the MRED) $S^*$. One can think of the distortion function, or equivalently the moment constraints, as reflecting attitudes towards risk. Distorted risk measures can therefore be tailored to specific risk attitudes through specification of the distortion function and its parameters.

In fact, for nonnegative random variables and differentiable distortion functions, distorted risk measures can be regarded as expected utility, for some implied utility function $u$ (Wirch and Hardy (1999)). In particular,

$$\rho_g(X) = \int_0^{\infty} g(S(x))dx = \int_0^{\infty} -xg'(S(x))f(x)dx = E[u(-X)], \tag{4.69}$$

where $u(y) = yg'(S(y))$, $y \leq 0$ is a utility function. This underlines the fact that distortion functions, or equivalently moment constraints in a relative entropy optimisation problem, reflect risk preferences. Note that the implied utility function $u$ depends on the distribution $S$ as well as the distortion function $g$.

Distorted risk measures also play a large role in two other economic theories. One of which is due to Yaari (1987), the other to Schmeidler (1989).

### 4.4.2 Examples

For the purposes of exposition, in all of these examples assume that $X$ is a continuous random variable and note that distorted risk measures are expectations taken with respect to the DPD, or equivalently, the MRED.

#### Example E1

In this example, the distortion function

$$g(u) = \min\left(\frac{(u-(1-\beta))^+}{\beta-\alpha}, 1\right) \quad \text{for} \quad 0 \leq \alpha < \beta \leq 1, u \in [0,1], \tag{4.70}$$

is concave if and only if $\beta = 1$. Thus by Corollary 4.11, the distorted risk measure $\rho_g$ is coherent only for $\beta = 1$. In this case $\rho_g$ is the risk measure known as the conditional tail expectation (CTE) or Tail-VaR defined as $E[X|X > F^{-1}(\alpha)]$. When $F^{-1}(\alpha)$ is subtracted from $\rho_g$ we obtain $\rho_g(X) - F^{-1}(\alpha) = E[X - F^{-1}(\alpha)|X > F^{-1}(\alpha)]$, which is the mean excess function of $X$.

If we take the limit as $\beta \downarrow \alpha$, the distortion function becomes a step function

with a step of height 1 at $\alpha$. In this case

$$\rho_g(X) = \inf\{x | P[X \leq x] \geq \alpha\}, \tag{4.71}$$

which is the Value-at-Risk or quantile reserve at the $100\alpha$th percentile. Value-at-Risk does not satisfy the subadditivity property and hence is not coherent (Artzner et al 1999).

**Example E2**

Here, the piece-wise linear distortion function

$$g(u) = \min\left((1+r)u, \frac{\alpha r + (1-\alpha(1+r))u}{1-\alpha}\right), \quad u \in [0,1] \tag{4.72}$$

for $0 < \alpha \leq 1, \quad -1 \leq r \leq \frac{1}{\alpha} - 1$ is concave if and only if $\alpha(1+r) \geq \alpha \Leftrightarrow r \geq 0$. Therefore by Corollary 4.11, the distorted risk measure $\rho_g$ is coherent only for $r \geq 0$. With $\alpha = 0.5$ and for $r \geq 0$, $\rho_g$ is Denneberg's absolute deviation principle (Wang (1996)).

**Example E3: Exponential Distortion**

In this case, the distortion function

$$g(u) = \frac{1 - e^{-\alpha u}}{1 - e^{-\alpha}}, \ \alpha > 0, \ u \in [0,1] \tag{4.73}$$

is always concave and is twice continuously differentiable. Hence, $\rho_g$ is a coherent risk measure (by Corollary 4.12).

### Example E4: Beta Distortion

In this example, the distortion function

$$g(u) = \int_0^u \frac{1}{\beta(a,b)} t^{a-1}(1-t)^{b-1}dt \quad a > 0,\ b > 0, \tag{4.74}$$

is concave if and only if $a \leq 1$ and $b \geq 1$. Furthermore, $g$ is twice continously differentiable. Therefore by Corollary 4.12, the beta-distorted risk measure $\rho_g$ is coherent if and only if $a \leq 1$ and $b \geq 1$. The parameters $a$ and $b$ control the effect of upper and lower tails respectively. The application of the beta transform to risk measurement was first proposed by Wirch (1999).

### Example E5: Proportional Hazards Distortion

The PH distortion is a special case of the beta distortion obtained when $b = 1$. From the above discussion, $\rho_g$ is coherent if and only if $a \leq 1$, which has the effect of inflating the right tail of the distribution of $X$. This distorted risk measure has been extensively studied in insurance applications (Wang (1995, 1996, 1996), Wang et al (1997)).

### Example E6: Dual Power Distortion

The dual power distortion is a special case of the beta distortion obtained by setting $a = 1$. Again, by the discussion of the beta distortion, $\rho_g$ is coherent if and only if $b \geq 1$. The parameter $b$ has the effect of deflating the lower tail of the distribution (Wang (1996)).

**Example E7**

**Logarithmic Distortion ($\gamma = 0$)**

The logarithmic distortion function

$$g(u) = \frac{\ln(a+bu) - \ln a}{\ln(a+b) - \ln a}, \quad a > 0,\ (a+b) > 0, \tag{4.75}$$

is concave if and only if $b \geq 0$. Furthermore, $g$ is twice continuously differentiable. Hence by Corollary 4.12, the distorted risk measure $\rho_g$ is coherent if and only if $b \geq 0$. When $a = 1$ and $b = r > 0$ this corresponds to the logarithmic transform given in Wang (1996).

**Power Distortion ($\gamma \neq 0$)**

In this case, the distortion function

$$g(u) = \frac{(a+bu)^\gamma - a^\gamma}{(a+b)^\gamma - a^\gamma}, \quad a > 0,\ (a+b) > 0,\ \gamma > 0, \tag{4.76}$$

is concave if and only if either ($b \geq 0$ and $\gamma \leq 1$) or ($-a < b \leq 0$ and $\gamma \geq 1$) which are exactly the conditions for the coherence of the corresponding distorted risk measure (Corollary 4.12). As stated earlier, $\gamma = 0.5$ gives a square root transform and $\gamma = 2$ gives a quadratic distortion, which also corresponds to the Gini principle in the study of income distributions (Wang (1996)).

**Example E8: Normal Distortion**

The normal distortion function is concave if and only if $\alpha \geq 0$. Since $g$ satisfies the conditions of Corollary 4.12, $\alpha \geq 0$ if and only if the corresponding distorted risk measure is coherent.

**Example E9**

Using only 1 moment constraint by setting $h_2 = 0$, the distorted risk measure $\rho_g$ takes the form of some common premium principles. Namely, $\rho_g$ is

1. the net premium principle when $h_1(u) = 1$,
2. a modified variance principle when $h_1(u) = \ln u$, and
3. the Esscher premium principle when $h_1(u) = u$ (Kamps (1998)).

We will not talk about coherence for these cases, as the distortion function depends very much on the ddf of $X$. When $h_1$ corresponds to a ddf, the resulting $\rho_g$ has a certain renewal-theoretic interpretation (Kamps (1998)).

## 4.5 Example: Exponential Family Distortions

The use of the normal distortion in Example E8 to price financial and insurance risks through the Choquet integral was originally proposed by Wang (1998). He shows that this transform has some desirable properties. In particular, this distortion is able to reproduce and generalise the Capital Asset Pricing Model (CAPM), to reproduce the Black-Scholes formula and provides a symmetric treatment of assets

and liabilities (due to the symmetric nature of the normal density function). Here, we propose a generalised distortion that includes the normal distortion as a special case and also provide the moment constraints in a relative entropy formulation of the DPD.

Let $F_e$ be the cdf of a random variable who is a member of the exponential family. Since $F_e$ is a member of the exponential family, its density function, $f_e$, can be written as

$$f_e(x) = \exp\left\{\sum \lambda_i h_i(x) - \psi(\lambda)\right\}, \tag{4.77}$$

for some functions $h_i, \psi$ and some constants $\lambda_i$. Now, define the distortion function

$$g(u) = F_e\left[F_e^{-1}(u) + \alpha\right], \tag{4.78}$$

where $F_e^{-1}$ is the inverse of $F_e$. The derivative is

$$\begin{aligned} g'(u) &= \frac{f_e(F_e^{-1}(u) + \alpha)}{f_e(F_e^{-1}(u))} \\ &= \exp\left\{\sum \lambda_i \left[h_i(F_e^{-1}(u) + \alpha) - h_i(F_e^{-1}(u))\right]\right\} \\ &= \exp\left\{\sum \lambda_i \tilde{h}_i(u) - \tilde{\psi}(\lambda)\right\}, \end{aligned} \tag{4.79}$$

where $\tilde{h}_i(u) = h_i(F_e^{-1}(u) + \alpha) - h_i(F_e^{-1}(u))$ and $\tilde{\psi}(\lambda) = 0$. Note that we constrain $F_e$ to be a member of the exponential family only so that it is possible to write $g'$ explicitly as an exponential function.

Let $X$ be a continuous random variable with ddf $S$. Since $g$ is differentiable,

Theorem 4.2 implies that the DPD $S^*$ is also an MRED where the set of moment constraints is determined by the distortion function $g$. In particular, this distribution satisfies

$$E_{S^*}\left[h_i(F_e^{-1}(S(X)) + \alpha) - h_i(F_e^{-1}(S(X)))\right] = 0 \quad \forall\, i. \tag{4.80}$$

The second derivative of $g$ is

$$g''(u) = \frac{\exp\{\sum \lambda_i \theta_i\}}{f_e(F_e^{-1}(u))} \left\{ \sum \lambda_i \left[h_i'(F_e^{-1}(u) + \alpha) - h_i'(F_e^{-1}(u))\right] \right\}. \tag{4.81}$$

where $\theta_i = (h_i(F_e^{-1}(u) + \alpha) - h_i(F_e^{-1}(u)))$. Theorem 4.3 implies that a necessary and sufficient condition for $g$ to be concave is

$$\sum \lambda_i \left[h_i'(F_e^{-1}(u) + \alpha) - h_i'(F_e^{-1}(u))\right] \leq 0. \tag{4.82}$$

Corollary 4.12 says that this is also a necessary and sufficient condition for the coherence of the corresponding distorted risk measure.

The normal transform corresponds to the case $F_e = \Phi$ where $\Phi$ is the $N(0,1)$ cdf. If we insist on symmetric treatment of assets and liabilities, then this can be accomplished with any symmetric distribution $F_e$.

### 4.5.1 Option Pricing via Normal Distortion

In finance, it is usual to value assets by computing expectations with respect to a suitable pricing measure. With a complete market model, the pricing measure is

unique and its construction is completely dictated by the market model (e.g., the Girsanov change of measure in the Black-Scholes market model). The completeness of the market model does not allow any flexibility in how one transforms or distorts the $P$ measure into a $Q$ measure. However, in an incomplete market model there may be a set of pricing measures and, to value an asset in this model, a method of selecting a pricing measure must be decided (e.g., relative entropy, utility function, distortion). In this case, there is some latitude in how one distorts the $P$ measure into a $Q$ measure.

In Example E8. we have shown that the risk-adjusted distribution (having density $f_N^*$) obtained from applying the normal distortion is also an MRED where relative entropy is minimised subject to the constraint

$$E_f[\Phi^{-1}(S(X))] = -\alpha. \tag{4.83}$$

A fundamental result in financial economics is that the underlying asset, properly discounted, is a martingale with respect to a pricing measure. This implies a constraint on the mean of the underlying asset under $Q$, namely

$$E_{f_Q}[X] = \gamma \tag{4.84}$$

for some constant $\gamma$, where $f_Q$ is a pdf in the set of acceptable pricing measures. This moment constraint results in an MRED having a density of the form

$$f_Q^*(x) = f(x)\exp(\lambda_Q x - \psi_Q(\lambda_Q)), \tag{4.85}$$

where $\psi'_Q(\lambda_Q) = \gamma$. It is interesting to investigate if there are conditions under which the two densities $f^*_N$ and $f^*_Q$ are the same.

**Lemma 4.13.** *The risk-adjusted densities $f^*_N$ and $f^*_Q$ are the same (i.e., $f^*_N(x) = f^*_Q(x)$ for all $x$) if and only if the original distribution is normal.*

Proof:

We have that

$$f^*_N(x) = f(x)\exp\left(-\alpha\Phi^{-1}(S(x)) - \frac{\alpha^2}{2}\right) \tag{4.86}$$

and

$$f^*_Q(x) = f(x)\exp\left(\lambda_Q x - \psi_Q(\lambda_Q)\right). \tag{4.87}$$

Thus $f^*_N(x) = f^*_Q(x)$ for all $x$ if and only if

$$-\alpha\Phi^{-1}(S(x)) - \frac{\alpha^2}{2} = \lambda_Q x - \psi_Q(\lambda_Q) \tag{4.88}$$

for all $x$. Solving for the ddf gives

$$S(x) = \Phi\left(\frac{\lambda_Q x - \psi_Q(\lambda_Q) + \frac{\alpha^2}{2}}{-\alpha}\right). \tag{4.89}$$

This is true for all $x$ if and only if $S$ is the ddf of a $N(\frac{-1}{\lambda_Q}(\psi_Q(\lambda_Q) - \frac{\alpha^2}{2}), (\frac{\alpha}{\lambda_Q})^2)$ random variable.

### 4.5.2 Normal Distortion from Generalised Relative Entropy Optimisation

In view of the Generalised Entropy Optimisation Principle in Remark 1.7, it is interesting to investigate whether a distribution distorted with the normal transform may be derived using a generalised measure of relative entropy, with the more natural constraint on $E_{fQ}[X]$.

Let us restrict ourselves to Csiszer's family of measures of directed divergence which, for continuous random variables is given by

$$H^C(\tilde{f}|f) = \int_\Omega f(x)\Psi\left(\frac{\tilde{f}(x)}{f(x)}\right)dx, \tag{4.90}$$

where $\Psi$ is a twice-differentiable convex function for which $\Psi(1) = 0$ and $\tilde{f}$ and $f$ are density functions.

Let $X$ be a continuous random variable with pdf $f$ and let $f_N^*$ be the dpdf computed using the normal transform. We wish to constrain the mean of $X$ under $f_N^*$. Then, the problem is to find a function $\Psi$ such that when $H^C$ is minimised using Lagrange's method subject to the constraint on the mean of $X$, the solution is given by $f_N^*$.

Following Section 7.2.2 in Kapur and Kesavan (1992), we see that $\Psi$ must satisfy

$$\Psi'\left(\frac{f_N^*(x)}{f(x)}\right) = \lambda_0 + \lambda_1 x \tag{4.91}$$

or

$$f_N^*(x) = f(x)\Psi'^{-1}(\lambda_0 + \lambda_1 x). \tag{4.92}$$

Furthermore, from our analysis of distortion functions, the likelihood ratio is

$$\frac{f_N^*(x)}{f(x)} = g'(S(x)) = \exp\left(-\alpha\Phi^{-1}(S(x)) - \frac{\alpha^2}{2}\right). \tag{4.93}$$

Using the change of variable

$$u = \exp\left(-\alpha\Phi^{-1}(S(x)) - \frac{\alpha^2}{2}\right) \tag{4.94}$$

and Equation 4.91 we see that

$$\Psi'(u) = \lambda_0 + \lambda_1 S^{-1}\left\{\Phi\left[\frac{\ln u + \frac{\alpha^2}{2}}{-\alpha}\right]\right\}. \tag{4.95}$$

Integrating this we get

$$\Psi(u) = \int_0^u \left(\lambda_0 + \lambda_1 S^{-1}\left\{\Phi\left[\frac{\ln z + \frac{\alpha^2}{2}}{-\alpha}\right]\right\}\right) dz, \tag{4.96}$$

and combining with the condition $\Psi(1) = 0$ gives

$$\Psi(u) = ku + \lambda_1 \int_0^u S^{-1}\left\{\Phi\left[\frac{\ln z + \frac{\alpha^2}{2}}{-\alpha}\right]\right\} dz, \tag{4.97}$$

where

$$k = -\lambda_1 \int_0^1 S^{-1} \left\{ \Phi \left[ \frac{\ln z + \frac{\alpha^2}{2}}{-\alpha} \right] \right\} dz. \tag{4.98}$$

Therefore, using the normal transform to compute a pricing measure is equivalent to computing the distribution that minimises the generalised relative entropy with respect to the original distribution, subject to the moment condition constraining the mean of the underlying asset. The generalised relative entropy is given by Csiszer's measure with $\Psi$ defined as above.

Using the special case $X \sim N(\mu, \sigma^2)$, it is easily shown that

$$\Psi(u) = \frac{\lambda_1 \sigma}{\alpha} u \ln u \tag{4.99}$$

giving

$$\begin{aligned} H^C(\tilde{f}|f) &= \int_\Omega f(x) \left( \frac{\lambda_1 \sigma}{\alpha} \frac{\tilde{f}(x)}{f(x)} \ln \left( \frac{\tilde{f}(x)}{f(x)} \right) \right) dx \\ &= \frac{\lambda_1 \sigma}{\alpha} \int_\Omega \tilde{f}(x) \ln \left( \frac{\tilde{f}(x)}{f(x)} \right) dx \end{aligned} \tag{4.100}$$

which is the usual measure of relative entropy. This result is another argument explaining why the normal distortion is able to reproduce the CAPM and the Black-Scholes formulae. In these cases, the logreturns are assumed to be normally distributed and constraints are placed on the mean returns (see Section 2.3 for a discussion on constraining the mean return versus the mean stock price).

## 4.6 Conclusion

We have established the link between relative entropy and distortion. Conditions on the distortion function or moment constraints that ensure the coherency of the corresponding distorted risk measure are provided. Furthermore, we have given results that allow one to verify whether a risk measure with a Choquet integral representation is coherent.

Risk preferences are contained in the distortion function or in the equivalent moment constraints in the relative entropy framework. The link to relative entropy provides additional intuition behind the many premium principles and risk measures having a Choquet integral representation. In addition, relative entropy provides an easy way of constructing new coherent risk measures by prescribing a new set of moment constraints. It may prove interesting to apply concepts from extreme value theory in the construction of risk measures. For example, a set of constraints may be imposed such that the MRED has the same mean excess function as the generalised Pareto distribution (GPD). The GPD is an integral tool in extreme value theory as it appears as the limit distribution of scaled excesses over high thresholds (Embrechts et al (1997)).

In financial economics, typical pricing measures impose a constraint on the mean of underlying asset. However, the normal distortion does not correspond to this moment constraint when working with the usual measure of relative entropy. To rectify this, a generalised measure of relative entropy is constructed such that, when minimised subject to a constraint on the mean, the same distribution as that obtained by the normal distortion is produced.

In closing, the connection between relative entropy and distortion gives a new perspective to the growing body of literature where the Choquet is applied to problems in finance, risk management and insurance.

# Chapter 5

# Open Questions and Directions for Future Work

There remain some open questions and directions for future work on the three topics presented in this thesis. For example, it may be interesting to investigate the sensitivity of the REB models to the historical data set. The inclusion of large crashes such as that observed in October of 1987 may have some impact on the performance of these models. However, we do not expect this impact to be very significant. The reweighted distribution with a prescribed mean and variance will inflate the probability of large gains, or deflate the probability of large losses, offsetting the impact of the observed large crashes on the pricing of options.

The results of Chapter 2 suggest that price discreteness is not a large factor in the pricing of S&P500 options. We performed a small study to further investigate the effect of discreteness. In particular, we rounded the historically observed prices to different accuracies and found that the rounding had little effect on the model

option prices. It is of interest to see if discreteness has a significant effect when pricing exotic options and/or options on individual stocks.

Further work on variance reduction is needed for the REB models to become efficient pricing tools. Also, it would be interesting to generalise the REB by developing a nonparametric bootstrap of volatility, possibly by using the observed squared returns. Another potential direction is to apply the REB models to the pricing of foreign exchange and fixed income securities. Stutzer and Chowdhury (1999) have extended Stutzer's approach to pricing bond futures options, providing the groundwork for future comparisons with the dynamic REB models.

It may be possible to extend the calibration algorithm of Chapter 3 to a dynamic version by using time series information on the underlying and/or set of options. In addition, by assuming some prior distribution for the parameters, a Bayesian adaptation of this algorithm may prove useful. Another open question is if the failure of models with constant parameters across maturities, such as that observed by the Hull-White model, is apparent when pricing options on a wide variety of individual stocks.

Further studies on the links between relative entropy, distortion and utility theory are needed to gain a deeper understanding of these relationships. The application of extreme value theory in the construction of risk measures using relative entropy is another path worthy of investigation. Finally, the exponential family distortions of Section 4.5 may be extended to include more general transformations.

# References

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, LIII(2):499-547.

Amin, K. I. and Ng, V. K. (1993). Option valuation with systematic stochastic volatility. *Journal of Finance*, XLVIII(3):881-910.

Andersen, L. and Andreasen, J. (1999). Jump-diffusion processes: Volatility smile fitting and numerical methods for pricing. Working paper, General Re Financial Products.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203-228.

Avellaneda, M. (1998). Minimum-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance*, 1(4):447-472.

Avellaneda, M., Buff, R., Friedman, C., Grandchamp, N., Kruk, L., and Newman, J. (2001). Weighted Monte Carlo: A new technique for calibrating asset-pricing models. *International Journal of Theoretical and Applied Finance*, 4(1):91-119.

Avellaneda, M., Friedman, C., Holmes, R., and Samperi, D. (1997). Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance*, 4(1):37-64.

Avellaneda, M. and Gamba, R. (2000). Conquering the greeks in Monte Carlo: Efficient calculation of the market sensitivities and hedge-ratios of financial assets by direct numerical simulation. Working paper, Courant Institute of Mathematical Sciences, New York University, New York.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, LII(5):2003–2049.

Benveniste, A., Métivier, M., and Priouret, P. (1987). *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, New York.

Berliner and Lev (1980). On the use of maximum entropy concept. *Transactions of the 21st ICA*, 1:47–62.

Bick, A. (1982). Comments on the valuation of derivative assets. *Journal of Financial Economics*. 10:331–345.

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, New York, third edition.

Björk, T. (1998). *Arbitrage Theory in Continuous Time*. Oxford University Press Inc., New York.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskdasticity. *Journal of Econometrics*, 31:307–327.

Box, G. E. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces.* John Wiley & Sons, New York.

Box, G. E., Hunter, W. G., and Hunter, S. J. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building.* John Wiley & Sons, New York.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621-651.

Britten-Jones, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *Journal of Finance*, 55(2):839-866.

Buchen, P. W. and Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31(1):143-159.

Bühlmann, H., Delbaen, F., Embrechts, P., and Shiryaev, A. N. (1996). No-arbitrage, change of measure and conditional Esscher transforms. *C.W.I. Quarterly*, 9(4):291-317.

Campbell, J. Y., Lo, A. W., and Mackinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, Princeton, New Jersey.

Carr, P., Geman, H., and Madan, D. B. (2000). Pricing and hedging in incomplete markets. *Journal of Financial Economics*, forthcoming.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2000). The fine structure of

asset returns: An empirical investigation. Working paper, Maryland Business School.

Chateauneuf, A., Kast, R., and Lapied, A. (1996). Choquet pricing for financial markets with frictions. *Mathematical Finance*, 6(3):323–330.

Chernov, M. and Ghysels, E. (1998). What data should be used to price options? Technical Report 98s–22, CIRANO, Montreal.

Christensen, B. J. and Kiefer, N. M. (2000). Simulated moment methods for empirical equivalent martingale measures. In Mariano, R., Schuermann, T., and Weeks, M. J., editors, *Simulation-based Inference in Econometrics*, chapter 7, pages 183–204. Cambridge University Press, Cambridge, U.K.

Clewlow, L. and Carverhill, A. (1995). *Over the Rainbow*, chapter Quicker on the Curves, pages 81-83. Risk Publications, London.

Colwell, D. B. and Elliott, R. J. (1993). Discontinuous asset prices and non-attainable contingent claims. *Mathematical Finance*, 3(3):295–308.

Cover, T. and Thomas, J. (1991). *Elements of Information Theory*. John Wiley & Sons, New York.

Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145-166.

Denneberg, D. (1994). *Non-Additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Derman, E. and Kani, I. (1994). Riding on a smile. *RISK*, 7:32-39.

Derman, E. and Kani, I. (1998). Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility. *International Journal of Theoretical and Applied Finance*, 1:7-22.

Duan, J.-C. (1995). The GARCH option pricing model. *Mathematical Finance*, 5(1):13-32.

Duan, J.-C. (1997). Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics*, 79:97-127.

Duan, J.-C. and So, M. K. (1999). Testing conditional ditributions in dynamic models using relative entropy. Technical report, Hong Kong University of Science and Technology.

Duffie, D. (1996). *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, New Jersey, 2nd edition.

Duffie, D. and Singleton, K. J. (1993). Simulated moments estimation of markov models of asset prices. *Econometrica*, 61(4):929-952.

Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *Journal of Finance*, LIII(6):2059-2106.

Dupire, B. (1994). Pricing with a smile. *RISK*, 7:18-20.

Efron, B. (1983). *The Jackknife, the Bootstrap and Other Resampling Plans*. J.W. Arrowsmith Ltd., Bristol, England.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall Ltd., New York.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Milan.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variances of U.K. inflation. *Econometrica*, 50:987–1008.

Ermoliev, Y. and Wets, R. J.-B., editors (1988). *Numerical Techniques for Stochastic Optimization.* Springer-Verlag, New York.

Föllmer, H. and Leukert, P. (1998). Quantile hedging. Preprint, Humboldt University, Berlin.

Föllmer, H. and Schweizer, M. (1991). Hedging of contingent claims under incomplete information. In Davis, M. and Elliott, R., editors, *Applied Stochastic Analysis*, volume 5 of *Stochastics Monographs.* Gordon and Breach, London.

Föllmer, H. and Sondermann, D. (1986). Hedging of non-redundant contingent claims. In Hildenbrand, W. and Mas-Colell, A., editors, *Contributions to Mathematical Economics.* Elsevier Science Publishers B.V., North-Holland.

Frittelli, M. (2000a). Introduction to a theory of value coherent with the no-arbitrage principle. *Finance and Stochastics*, 4:275–297.

Frittelli, M. (2000b). The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10(1):39–52.

Gallant, R. A. and Tauchen, G. (1995). Which moments to match? *Econometric Theory*, 12(4):657–681.

Gerber, H. and Shiu, E. (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries*, 46:99–140.

Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118.

Gouriéroux, C., Renault, E., and Touzi, N. (2000). Calibration by simulation for small sample bias correction. In Mariano, R., Schuermann, T., and Weeks, M. J., editors, *Simulation-based Inference in Econometrics*, chapter 13, pages 328–358. Cambridge University Press, Cambridge, U.K.

Gulko, B. L. (1998). *The Entropy Pricing Theory*. PhD thesis, Yale University.

Hardy, M. R. (1999). A regime switching model of long term stock returns. Research Report 99-04, Institute of Insurance and Pension Research, University of Waterloo, Waterloo, Ontario, Canada.

Hasbrouck, J. (1999). The dynamics of discrete bid and ask quotes. *Journal of Finance*.

Heston, S. L. (1993a). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Heston, S. L. (1993b). Invisible parameters in option prices. *Journal of Finance*, XLVIII(3):933–947.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, XLII(2):281–300.

Hull, J. C. (1997). *Options, Futures, and Other Derivatives*. Prentice-Hall, Inc., Upper Saddle River, NJ, third edition.

Jackwerth, J. C. (1999). Option-implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives*, 7(2):66–82.

Jackwerth, J. C. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *Journal of Finance*, LI(5):1611–1631.

Jarrow, R. and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10:347–369.

Kamps, U. (1998). On a class of premium principles including the Esscher principle. *Scandinavian Actuarial Journal*, 1:75–80.

Kapur, J. and Kesavan, H. (1992). *Entropy Optimization Principles with Applications*. Academic Press, Inc., San Diego, CA.

Karatzas, I., Lehoczky, J., Shreve, S., and Xu, G. (1991). Martingale and duality methods for utility maximization in an incomplete market. *SIAM Journal of Control and Optimization*, 29:702–730.

Karatzas, I. and Shreve, S. E. (1998). *Methods of Mathematical Finance*. Springer-Verlag, New York.

Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874.

Laurent, J.-P. and Leisen, D. P. (1999). Building a consistent pricing model from observed option prices. Working paper, the Hoover Institute, Stanford University.

Ledoit, O. and Santa-Clara, P. (1998). Relative pricing of options with stochastic volatility. Technical Report 09–98, UCLA.

Maeder (1982). Asymetrie de la distribution collective des sinistres et probabilite de ruine. *BASA*, pages 91–115.

Martin, V. L. and Pagan, A. R. (2000). Simulation-based estimation of some factor models in econometrics. In Mariano, R., Schuermann, T., and Weeks, M. J., editors, *Simulation-based Inference in Econometrics*, chapter 10, pages 235–254. Cambridge University Press, Cambridge, U.K.

McFadden, C. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57:995–1026.

McLeish, D. L. (1998). Some simple properties of high, low, open, close: Simulating financial time series and tracking volatility. Technical Report 98-04, University of Waterloo.

Melick, W. R. and Thomas, C. P. (1997). Recovering an asset's implied PDF from option prices: An application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis*, 32:91–115.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144.

Montgomery, D. C. (1997). *Design and Analysis of Experiments*. John Wiley & Sons, New York, fourth edition.

Pakes, A. and Pollard, D. (1989). The asymptotics of simulation estimators. *Econometrica*, 57:1027–1058.

Panjer, H. H., editor (1998). *Financial Economics: With Applications to Investments, Insurance and Pensions*. The Actuarial Foundation, Schaumburg, Illinois.

Pastorello, S., Renault, E., and Touzi, N. (2000). Statistical inference for random-variance option pricing. *Journal of Business and Economic Statistics*, 18(3):358–367.

Redekop, J. (1995). *Extreme-Value Distributions for Generalizations of Brownian Motion*. PhD thesis, University of Waterloo.

Ritchken, P. and Trevor, R. (1999). Pricing options under generalized GARCH and stochastic volatility processes. *Journal of Finance*, LIV(1):377–402.

Rodgers, L. (1994). Equivalent martingale measures and no-arbitrage. *Stochastics and Stochastics Reports*, 51:41–49.

Rollans, S. and McLeish, D. L. (2002). Estimating the optimum of a stochastic system using simulation. *Journal of Statistical Computation and Simulation*, to appear.

Rouge, R. and Karoui, N. E. (2000). Pricing via utility maximization and entropy. *Mathematical Finance*, 10(2):259-276.

Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. *Journal of Finance*, XL(2):455-480.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance*, 49(3):771-818.

Rubinstein, M. (1998). Edgeworth binomial trees. *Journal of Derivatives*, 5(3):20-27.

Samperi, D. (1998). *Inverse Problems, Model Selection and Entropy in Derivative Security Pricing*. PhD thesis, New York University.

Schmeidler, C. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57:571-587.

Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatiltiy: An analytic approach. *The Review of Financial Studies*, 4(4):727-752.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *Journal of Finance*, LI(5):1633-1652.

Stutzer, M. and Chowdhury, M. (1999). A simple non-parametric approach to bond futures option pricing. *Journal of Fixed Income*, 8(4):67-76.

Sundt (1982). Minimum entropy in risk theory. *BASA*, pages 296-299.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17:43-54.

Wang, S. (1996a). Ordering of risks under ph-transforms. *Insurance: Mathematics and Economics*, 18:109-114.

Wang, S. (1996b). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(1):71-92.

Wang, S. and Dhaene, J. (1998). Comonotonicity, correlation order and premium principles. *Insurance: Mathematics and Economics*, 22:235-242.

Wang, S. S. (1999). A class of distortion operators for pricing financial and insurance risks. Research Report 99-08, Institute of Insurance and Pension Research, University of Waterloo, Waterloo, Ontario.

Wang, S. S. and Young, V. R. (1998). Risk-adjusted credibility premiums using distorted probabilities. *Scandinavian Actuarial Journal*, 2:143-165.

Wang, S. S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21:173-183.

Welch, W. J., Yu, T.-K., Kang, S. M., and Sacks, J. (1990). Computer experiments for quality control by parameter design. *Journal of Quality Technology*, 22(1):15-22.

Wiggins, J. B. (1987). Option values under stochastic volatility. *Journal of Financial Economics*, 19:351–372.

Wirch, J. L. (1999). *Coherent Beta Risk Measures for Capital Requirements*. PhD thesis, University of Waterloo, Waterloo, Ontario.

Wirch, J. L. and Hardy, M. R. (1999). A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, 25:337–347.

Wirch, J. L. and Hardy, M. R. (2000). Distortion risk measures: Coherence and stochastic dominance. Working paper presented to 2000 AFIR Colloquium, Tromso, Norway.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55(1):95–115.